Selection Bias and Racial Disparities in Police Use of Force*

Felipe Goncalves[†] Steven Mello[‡] Emily Weisburst[§]

August 18, 2025

Abstract

We study racial disparities in police use of force. A pervasive issue in studies of policing is that the available data are selected by the police. As a result, disparities computed in the observed sample may be biased if selection into the data differs by race. We develop a framework and econometric strategy for correcting this bias, using variation across officers in enforcement intensity to identify the racial composition of the unobserved population at risk of selection. Using detailed administrative data on arrests and force incidents from Chicago and Seattle, we find that Black civilians comprise 56 percent of arrestees but about 49 percent of potential arrestees. Correcting for sample selection doubles our measure of the racial disparity in force rates. Decompositions of the corrected force disparity reveal that about 70 percent is unexplained by other demographic and incident characteristics, suggesting an important role for officer discrimination. Our selection bias estimates meaningfully impact the conclusions drawn in the existing literature.

JEL Codes: J15, K42, C10

*We are grateful to David Arnold, Bocar Ba, Andrew Jordan, Patrick Kline, Adriana Lleras-Muney, Erzo Luttmer, Jonathan Mummolo, Susan Parker, Will Rafey, Roman Rivera, Andres Santos, Doug Staiger, CarlyWill Sloan, Yotam Shem-Tov, Bernardo Silveira, Jonathan Tebes, Till von Wachter, Heidi Williams, and Crystal Yang, as well as seminar participants at APPAM, Dartmouth, Princeton, TxECW, UC-Berkeley, University of South Carolina, the WashU Criminal Justice Conference, and the NBER Summer Institute for helpful comments. TJ Hedin provided excellent research assistance. Generous funding was provided by the UCLA California Center for Population Research and the Rockefeller Center at Dartmouth College.

[†]University of California, Los Angeles and NBER; fgoncalves@ucla.edu

[‡]Dartmouth College and NBER; steve.mello@dartmouth.edu

[§]University of California, Los Angeles and NBER; weisburst@ucla.edu

1 Introduction

The ability to use physical force against civilians is arguably the most grave exercise of power available to law enforcement. In recent years, high-profile police killings of unarmed Black civilians have brought racial disparities in police use of force to the forefront of public consciousness. Even prior to the killing of George Floyd and the associated mass protests during the summer of 2020, just 35 percent of Black civilians agreed that police treat racial groups equally and just 33 percent viewed the police as using appropriate levels of force (Desilver et al., 2020). Moreover, the disparate incidence of police force can have important implications for youth outcomes in minority communities (Ang, 2021) and for public cooperation with the police (Ang et al. 2025; Zaiour and Mikdash 2024). While unique in its severity, use of force is just one of many criminal justice outcomes in the United States marked by racial disparities. Black individuals are also more likely to be stopped by the police, convicted of a crime, denied bail, and issued a lengthy prison sentence than observably similar white individuals (National Academies of Sciences et al., 2022).

An important obstacle to understanding racial disparities in criminal justice outcomes arises from the fact that available data typically only include interactions initiated at the discretion of criminal justice agents. Researchers interested in studying bail decisions, criminal convictions, or incarceration, for example, are generally only able to observe these outcomes among a sample of individuals who police choose to arrest, prosecutors choose to charge with an offense, or judges choose to convict. The behavior of these agents may generate racial disparities on the margin of whom is selected into the data, rendering disparities measured in the selected sample challenging to interpret (e.g., Knox et al. 2020). In particular, disparities in the selected sample are often uninformative about the differential risk of treatment without additional strong and untestable assumptions.

In this paper, we develop a novel empirical strategy for estimating disparities when the data suffer from this type of sample selection bias, and we apply this strategy to study racial disparities in police use of force among arrestees. Our strategy combines two key insights: 1) a selection-corrected disparity can be calculated with an estimate of the racial composition of the potentially-selected sample; and 2) this racial composition can be identified from the agents with the highest propensity to select people into the sample. Using data from two large, urban police departments, we find strong evidence of racial disparities in arrests among potential arrestees. Accordingly, we find that disparities in police use of force which account for this selection bias are about twice as large as the naive disparity in the selected sample.

We begin by presenting a general conceptual framework where agents randomly encounter individuals and face two sequential, binary choices: whether to select an individual into the data and then whether to treat them. In our empirical application, these are a police officer's choices to (i) make an arrest and (ii) use force during the arrest. This conceptual framework highlights that racial disparities computed in the selected sample suffer from selection bias when there are racial differences in the probability of being selected into the data. It also illustrates that this sample selection bias can be corrected with an estimate of the racial composition of the potentially-selected sample of interest.

We then apply the ideas from our conceptual framework using detailed data from the Chicago and Seattle police departments, where we can observe officers' work shifts, arrest records with offense and arrestee information, and each incident of police use of force. We propose an econometric approach for estimating the key parameter for selection correction, the racial composition of the potentially-selected sample, which leverages variation across police officers in the propensity to select individuals into the data. The first step of our approach is to measure variation across officers in their enforcement intensity by estimating each officer's average number of arrests per shift they work, adjusting for detailed assignment fixed effects. We then use the racial composition of individuals arrested by the most-enforcing officers, again adjusting for detailed assignment effects, as our estimate of the racial composition of the potentially-arrested sample.

Our econometric approach relies on two assumptions. The first is an exogeneity assumption requiring that officers patrolling the same assignment encounter the same set of potential arrests. The second, which we term extremum-agent monotonicity, requires the existence of some maximally-enforcing officer such that, if any other officer arrests a particular civilian, this extremum officer does as well. In other words, there is some officer that makes any arrest which would also be made by another officer. These two assumptions jointly imply bounds on the composition of each officer's selected sample of arrestees, with these bounds determined by the composition of the extremum officer's arrests and a given officer's enforcement intensity relative to that of the extremum officer. Using these bounds, we can validate our assumptions by jointly testing a series a moment inequalities, and this test fails to reject the joint null of exogeneity and extremum officer monotonicity.

Given these assumptions, our approach recovers the racial composition of a particular sample of interest: individuals who would be arrested by some officer in our data. This is a conceptually and econometrically well-defined sample that represents the set of individuals who are actually at risk of facing force during an arrest. Considering the emphasis of U.S. case law on the application of "reasonable officer" standards, we also view this sample as likely representative of the set of individuals who are legally eligible to be arrested in practice, as we discuss further in section 2.3 below.

While Black civilians comprise 56 percent of the selected sample of arrestees, we estimate that they make up about 49 percent of the unselected sample of potential arrestees, indicating that differential selection into the data is empirically relevant in this setting.[1] Our estimates imply that among potential arrestees, Black individuals are about three percentage points, or thirty percent, more likely to be arrested. This estimate advances a large literature on racial disparities in the criminal justice system. In particular, while scholars have documented racial disparities in traffic and pedestrian stops and searches, as well as at various downstream stages of the system such as pretrial detention and sentencing, disparities at the arrest stage have received minimal attention.

Following from our conceptual framework, we then use our estimate of the racial composition of potential arrestees to compute an unselected (i.e., adjusted for differential sample selection) disparity in police use of force. In the selected sample, non-Black civilians face force in 2.3 percent of arrests and Black civilians are 0.55 percentage points more likely to experience force. Our estimate of the unselected disparity is 1.1 percentage points, or about twice as large as the naive disparity in the selected sample. The difference between selected and unselected disparities is statistically significant at conventional levels and this finding withstands an array of alternative specifications and various robustness tests.

While our baseline estimate corresponds to the racial disparity in risk of force among potential arrestees, it does not necessarily align with conventional notions of "causal" discrimination because this disparity could arise due to differences in non-race characteristics across groups. To contextualize our analysis within the broader discrimination literature, we use a Kitagawa-Oaxaca-Blinder (KOB) decomposition approach to parse the unselected disparity into components which are within and across non-race observables (gender, age, and criminal offense type). We find that about 70 percent of the unselected disparity is within-observables. This within-characteristics disparity corresponds to the causal effect of race on police use of force among potential arrestees under an additional selection on observables assumption, and we interpret the results of this KOB decomposition as suggesting an important role for racial discrimination in driving the disparity we find.

We also consider the implications of our findings for the conclusions in Fryer Jr (2019), who examines racial disparities in police use of force across a range of data sources and force outcomes. Notably, while Fryer Jr (2019) finds typical Black-white disparities for less severe forms of force, he finds, if anything, negative gaps (i.e., "reverse discrimination") when examining the most serious force types. Concerns about selection bias in Fryer Jr

[1] More specifically, we find that the share Black among an officer's arrestees declines with an officer's enforcement intensity and is the smallest, on average, among the most-enforcing officers.

(2019)'s seminal paper have prompted responses focused on the potential importance of unrepresentative samples (Durlauf and Heckman, 2020) or constructing bounds based on assumptions about selection bias (Knox et al., 2020), as well as motivated studies focusing on related but alternative estimands of interest (Ba et al., 2021; Hoekstra and Sloan, 2022; Schwartz and Jahn, 2020). Assuming a proportionally similar rate of differential selection into the data by race (within non-race observables) in his settings and ours, we compute selection-adjusted versions of the estimates in Fryer Jr (2019). We show that our estimated differential selection is sufficient to erode, but not reverse, his force gaps in officer-involved shootings. Our paper advances the emerging literature on racial disparities in police use of force in economics by providing the first estimates of racial disparities in police use of force which correct for sample selection bias.

Our methodological approach, which uses variation in enforcement intensity across police officers, builds on a growing literature documenting the importance of variation in discretionary behavior among criminal justice agents (e.g. Ba et al., 2021; Adger et al., 2022; Feigenberg and Miller, 2022; Chalfin and Goncalves, 2023; Goncalves and Mello, 2021, 2023; Weisburst, 2024; Rivera, 2025). Our approach is an application of "identification at infinity" ideas in selection models, originating with Chamberlain (1986) and Heckman (1990) and popularized by Hull (2020) and Arnold et al. (2022). We illustrate how these ideas can be applied to a broader set of questions and institutional contexts with our version of a monotonicity assumption.

Our paper also builds on a methodological literature in economics focusing on addressing identification and estimation issues arising from selection bias (e.g. Heckman, 1979; Manski, 1990; Lee, 2009; Dutz et al., 2021). Importantly, this literature typically focuses on contexts where the full population of interest is directly observed but the outcome data are incomplete; a notable example is studies of earnings inequality that account for non-employment (e.g. Neal, 2004; Blundell et al., 2007). In contrast, we provide a framework and empirical strategy for addressing selection bias when the population of interest is *unobserved*. Our methodological approach could be applied in a wide range of settings where agents have discretion in determining selection into administrative datasets, such as medical referrals, school admissions, or caseworker decisions in a wide array of social service contexts.

Finally, our paper contributes to an emerging literature on discrimination in systems with multiple stages (Baron et al. 2024; Bohren et al. 2025) by illustrating the relevance of racial disparities at one stage of the criminal justice system (arrests) for the inferences one can draw at later stages. Along these lines, we view our finding of differential arrest rates by race as an important contribution to the broader literature focused on understanding racial disparities in the criminal justice system. As shown in figure 1, evidence from nationally representative datasets suggests that arrests are the stage of the criminal justice system at which racial disparities meaningfully emerge. Moreover, while our analysis focuses on quantifying disparities in the risk of facing police use of force, the disparate arrest rates we document have important implications for understanding disparities at any stage which is downstream from an arrest. As an illustrative example, we provide a back-of-the-envelope calculation of the selection-adjusted disparity in pretrial detention rates using summary statistics reported in Dobbie et al. (2018). The extent of differential selection in our analysis implies a selection-adjusted disparity on this margin which is three times larger than the disparity in the selected sample of arrestees.

The rest of our paper is organized as follows. Section 2 presents our conceptual framework, highlighting the identification problem, interpreting our selection-corrected estimand, and defining the unobserved population of interest. In section 3, we discuss our data and empirical setting, providing additional context for our focus on arrests as the selection stage of interest. Section 4 lays out our empirical strategy for identifying the racial composition of the potentially-selected sample using the most enforcing officers and discusses the assumptions required for our approach. In section 5, we present our main results, various robustness checks, and tests of our key assumptions. Section 6 discusses and contextualizes our results in the broader literature on discrimination and disparities in police of force.

2 Conceptual framework

2.1 Identification problem

We consider a setting with individuals i who encounter a set of agents $j \in \mathcal{J}$. In each possible interaction, agents have two binary choices. First, they decide whether to select the individual into the sample, $S_{ij} \in \{0, 1\}$. In our analyses, this is an officer's choice to make an arrest but could represent, for example, the choice to make a traffic or pedestrian stop in other settings. Next, for those individuals with $S_{ij} = 1$, the agent decides whether to treat the individual, $D_{ij} \in \{0, 1\}$. In our setting, D_{ij} is the decision to use force during an arrest. We denote the officer who encounters individual i by j_i. The *realized* outcomes of an individual are $S_i = \sum_{j \in \mathcal{J}} \mathbb{I}[j_i = j] S_{ij}$ and $D_i = \sum_{j \in \mathcal{J}} \mathbb{I}[j_i = j] D_{ij}$.[2]

[2]Note that this setup is consistent with supposing that there is a potential treatment outcome D^*_{ij} in an interaction, with the realized treatment outcome depending on selection, $D_{ij} = D^*_{ij} S_{ij}$. We consider this extended formulation below when differentiating our estimand of interest from the target estimand in a traditional Heckit selection correction.

Individuals i differ by their race, $R_i \in \{b, w\}$,[3], other observable characteristics X_i, and unobservable characteristics θ_i. For now, we abstract away from non-race characteristics and focus our interest on an average racial disparity in treatment.

We are interested in individuals who comprise a particular sample, whose membership is defined by $P_i \in \{0, 1\}$. Our primary object of interest, which we denote by Δ, is the racial disparity in realized treatment among individuals in this sample:

$$\Delta = E[D_i|R_i = b, P_i = 1] - E[D_i|R_i = w, P_i = 1] \tag{1}$$

We refer to the group with $P_i = 1$ as the *target sample*. We keep this notion general for now, imposing only that this sample contains the set of individuals who would be selected by any agent, $P_i \geq \max_{j \in \mathcal{J}} S_{ij} \geq S_i$, and discuss the specific target sample of interest for our empirical application below in section 2.3. For notational simplicity, we suppress the conditioning on target sample membership throughout the remainder of this section and treat P_i as including all individuals in our setting.

Note that Δ does not necessarily correspond to a causal difference in treatment attributable to race. Specifically, $\Delta \neq 0$ could reflect disparate treatment attributable to non-race characteristics which are correlated with race. We begin by focusing on this unconditional estimand to highlight that sample selection bias presents issues even for estimating an unconditional disparity. Our first goal is to address the sample selection bias hurdle to identifying Δ. We then discuss the relationship between Δ and other notions of discrimination (i.e., disparities conditional on non-race characteristics), as well as compute versions of the conditional disparity, in section 6.1.

The main identification challenge arises because Δ cannot be estimated with an empirical analogue. We can observe individuals who are selected into the sample ($S_i = 1$) but not those who are unselected ($S_i = 0$). Comparing treatment outcomes only among those who are selected, which we term the selected disparity Δ_s, gives:

$$\begin{aligned}
\Delta_s &\equiv E[D_i|R_i = b, S_i = 1] - E[D_i|R_i = w, S_i = 1] \\
&= \frac{E[D_i|R_i = b]}{E[S_i|R_i = b]} - \frac{E[D_i|R_i = w]}{E[S_i|R_i = w]} \pm \frac{E[D_i|R_i = w]}{E[S_i|R_i = b]} \\
&= \frac{\Delta}{E[S_i|R_i = b]} + \left[\frac{E[S_i|R_i = w]}{E[S_i|R_i = b]} - 1\right] E[D_i|R_i = w, S_i = 1].
\end{aligned} \tag{2}$$

[3]Throughout our analysis, we define racial groups as Black (denoted by b) and non-Black (denoted by w). We use this definition because (i) our empirical approach requires a binary notion of race; (ii) Black versus non-Black disparities are particularly salient in public datasets; (iii) Hispanic status is not reported for arrestees in the data from one of our settings (Seattle).

The second equation follows from the fact that $S_i = 0$ implies $D_i = 0$. Equation (2) highlights that, for example, a relatively higher likelihood of selection for Black individuals ($E[S_i|R_i = b] > E[S_i|R_i = w]$) will yield a downward bias in the selected disparity, $\Delta_s < \Delta/E[S_i|R_i = b]$. It also highlights that Δ and Δ_s can even have different signs because they differ beyond a scaling factor.[4]

We can also observe here that, even without a sample selection issue, the parameters Δ_s and Δ have different scales, since the former is conditional on selection into the sample while the latter is conditional only on target sample membership. In most settings, we expect the selection rate to be quite low (e.g., in the U.S., only 12 percent of property crimes are cleared via an arrest). Hence, to identify an unselected disparity which is comparable to the selected disparity in terms of scale, we focus on the problem of identifying a rescaled version of Δ:

$$\tilde{\Delta} \equiv \Delta/E[S_i|R_i = b] \tag{3}$$

2.2 Solution to identification problem

Our goal is to estimate a disparity in treatment which is unconditional on sample selection. Using Bayes' rule to rewrite the race-specific selection probabilities in (2) yields the following expression for $\tilde{\Delta}$:

$$\tilde{\Delta} = \Delta_s + \left[1 - \frac{\pi}{1-\pi} \cdot \frac{1-\pi_s}{\pi_s}\right] E[D_i|R_i = w, S_i = 1], \tag{4}$$

where π and π_s are the Black shares of the target and selected samples, respectively. The selected data directly provide us with Δ_s, π_s, and $E[D_i|R_i = w, S_i = 1]$. Hence, the unselected race share π is the only unknown in (4), and the challenge of identifying $\tilde{\Delta}$ reduces to identifying the Black share of the target population.[5]

2.3 Target sample definition

The discussion above illustrates how one can adjust the disparity in a selected sample for selection bias arising due to differential selection into the data with an estimate of the racial composition of the target sample of interest. Of course, both the interpretation and empirical implementation of this idea depend critically on the notion of the target sample.

[4] In appendix figure A-1, we provide an intuitive and straightforward visualization of the identification problem highlighted by equation (2).

[5] Note that estimating the unscaled Δ (rather than the rescaled $\tilde{\Delta}$) also relies only on observed data plus an estimate of π and $E[S_i]$, since $\Delta = \tilde{\Delta} E[S_i|R_i = m] = \tilde{\Delta}\frac{\pi_s}{\pi} E[S_i]$.

In our empirical work, we examine use of force (D_i) occurring during arrests (S_i) and define the target sample of interest as:

$$P_i = \max_{j \in \mathcal{J}} S_{ij},$$

or in words, the set of individuals for whom there is at least one officer $j \in \mathcal{J}$ that would arrest them. This is a conceptually and econometrically well-defined sample and represents the set of individuals who are at risk of facing force during an arrest. As described further in section 4 below, we identify the racial composition of this sample from the racial composition of the most enforcing officers, or those with the highest propensity to make arrests.

Given the relevant case law and the *de facto*, if not *de jure*, legal environment with respect to the validity of officer arrest decisions, there is also an argument for viewing this sample as equivalent to the set of individuals who are *legally eligible to be arrested*. In the U.S., the legal standard for an arrest is a reasonable basis for believing that a crime may have been committed at the time of an arrest. Although charges may be dismissed by a prosecutor or judge after the fact, the legality of initial arrest decisions are rarely overturned. The Supreme Court has held that officers are protected against civil lawsuits under qualified immunity as long as another "reasonable" officer would have made a similar decision in comparable circumstances.[6]

While not necessarily given a formal or explicit treatment in the existing literature, implicit conceptual and empirical questions about the target sample (as we define it) lurk throughout several strands of research on the criminal justice system. For example, scholars interested in racial disparities in traffic or pedestrian stops have compared the racial composition of stops with that of various benchmarks, including stops during darkness (Grogger and Ridgeway, 2006), motorists involved in accidents (Alpert et al., 2004), stops made by specific officer subgroups (e.g., Ridgeway and MacDonald 2009; Ba et al. 2021), and benchmarks constructed from telematics data (e.g., Cai et al. 2022; Aggarwal et al. 2025), each of which implicitly corresponds to a choice of target sample.

The issue of the "correct" target sample has also proven salient in the courts and among policymakers. In response to arguments during *Floyd et al. vs. the City of New York* (2008), then-Mayor Bloomberg delivered a public critique of the plaintiff's claims that differences in the racial composition of pedestrian stops and the city's population reflect discriminatory

[6] See Legal Information Institute, Cornell University (https://www.law.cornell.edu/wex/probable_cause) and District of Columbia v. Wesby, 583 U.S. (2018) (https://supreme.justia.com/cases/federal/us/583/15-1485/). We provide some additional discussion of the relevant legal environment in appendix C.

behavior by the police, essentially arguing that the sample eligible to be stopped is different than the population as a whole (Bloomberg, 2013).

We acknowledge some potential caveats associated with our stance on the target sample. First, this definition is context-specific in the sense that it depends on the set of officers working; changes in the set of agents $j \in \mathcal{J}$ can correspond to changes in the set of individuals in the target sample.[7] And second, while this sample includes all individuals at risk of force during an arrest, it does not include all individuals at risk of force because some force incidents occur during interactions which do not result in an arrest.

We further discuss our focus on force among arrestees below in section 3.1 and we return to the question of the target sample in section 5.4. There, we report estimates when alternatively defining the full population as the target sample of interest, as well as discuss alternative target samples of interest beyond the one above and the potential of our empirics to speak to questions about these other target samples.

2.4 Interpreting our estimand

How do we interpret our target estimand $\tilde{\Delta}$, and how does it differ from alternative estimands for racial differences in use of force? This estimand measures the racial difference in the likelihood of experiencing force by individuals in the target sample. Note that this difference is an *observational* object, and our goal at first is not to interpret it as the causal effect of civilian race on police force. In section 6.1, we decompose Δ into components that are due to: 1) racial differences in other civilian demographics and incident characteristics, and 2) racial force gaps for individuals with the same characteristics. In appendix B, we show that this latter component can be interpreted as the causal effect of race on force for the target sample under a standard "conditional unconfoundedness" (or selection on observables) assumption.

Also worth highlighting here is how our estimand differs from that of standard selection correction approaches in labor economics *a la* Heckman (1979), since the distinction has important implications for both our conceptual framework and our empirical implementation. Suppose that there is a potential treatment outcome D^*_{ij} in an interaction which reflects whether force is used if the individual is selected into the sample, with the observed outcome given by $D_{ij} = D^*_{ij} S_{ij}$. The goal of the Heckman (1979) approach is to estimate $\Delta^* =$

[7] A related implication is that our target sample definition will include discrimination that is practiced broadly by all officers when making selection decisions. In other words, the target sample may differ from a group deemed sufficiently guilty to be arrested by some objective measure (e.g., the target samples in Cai et al. 2022 or Aggarwal et al. 2025). We discuss this consideration further in section 5.4 and provide evidence against the concern that the target sample we identify is "contaminated" by discrimination at the selection margin practiced by all officers.

$E[D^*_{ij}|R_i = b, P_i = 1] - E[D^*_{ij}|R_i = w, P_i = 1]$. In words, this estimand measures the force disparity *if all potential arrestees were arrested*, whereas our estimand measures the force disparity at observed values of arrest and force outcomes. In many applications of the Heckit approach, such as missing wages among the unemployed, this "fully selected" counterfactual is a much less dramatic extrapolation from the observed data than in our context, where we expect selection rates to be quite low (as noted above).

Although there are perhaps conceptual reasons to prefer one estimand over the other and vice versa, our estimand can be recovered under much weaker assumptions, as we discuss further in section 4 and appendix B.5. Implementation of the Heckit estimator requires the availability of an instrument that satisfies: 1) *strict monotonicity*, so that the instrument weakly shifts all individuals' selection outcome (arrest) in the same direction, and 2) *exclusion*, so that it impacts selection into the sample (arrest) but does not impact treatment (force). If officer identity were used in this estimator, the latter assumption requires that officer arrest propensity is independent of force propensity. In contrast, our empirical strategy, which we describe below, will impose neither strict monotonicity nor exclusion.

The existing literature has recognized the threat of sample selection bias to the study of racial disparities in the criminal justice system and our conceptual framework provides a useful structure for interpreting the various approaches taken in prior work. West (2018) studies a setting with no officer discretion over whether an incident enters the data, police investigations of car accidents, so that racial gaps are free of selection bias. In our framework, this institutional feature implies $S_i = P_i$. Knox et al. (2020), who develop a similar conceptual framework to ours, show how directional assumptions on the patterns of sample selection (e.g. $\pi_s \geq \pi$) can provide informative bounds on the selection-corrected force gap. Aggarwal et al. (2025) tackle this challenge in traffic stops data and use rich administrative data from a rideshare app to directly measure the distribution of true offending by race, allowing them to measure $E[S_i|R_i]$ and test for differences in being stopped by an officer by driver race. Another class of studies has relied on information about *officer race* and asks how force disparities by civilian race vary by officer race (e.g., Ba et al. 2021; Hoekstra and Sloan 2022; Antonovics and Knight 2009). While this strategy identifies a different target estimand from ours (Anwar and Fang, 2006), it offers an approach that is not susceptible to sample selection bias.

The closest paper to ours in terms of empirical setting and estimand of interest is Fryer Jr (2019), who estimates racial differences in police use of force using multiple datasets on individuals who encounter the police in the context of stops or arrests. Fryer Jr (2019) explicitly notes the concern of sample selection bias and takes the approach of conditioning on a rich set of observable characteristics X measuring a civilian's demographics and the circumstances of the police encounter. As we describe in section 6.2 below, this approach requires that $E[S_i|X_i, R_i] = E[S_i|X_i]$, or that race is uninformative about selection into the sample after conditioning on X.

3 Setting and data

3.1 Context

Approximately 2,000 people are killed by police each year in the U.S., and survey evidence suggests that 2 percent of police interactions (over 1 million incidents) involve the threat or use of force (Burghart, 2024; Tapp and Davis, 2022). While rare in absolute terms, police use of force is relatively common in the United States; police are responsible for 3.4 civilian deaths per 1 million persons in the U.S., a rate that is over 3 times higher than the rate in Canada and over 10 times higher than rates in England or France (Hirschfield, 2023). Research on police use of force in the U.S. has typically focused on fatal force events, as national data on these incidents is more readily available.

Our analysis focuses on the cities of Seattle, WA and Chicago, IL. Appendix table A-1 illustrates how these two cities to compare to other large and medium-sized cities in the U.S. using information from the FBI's Uniform Crime Reports (UCR), as well as information from the American Community Survey (ACS) and *Fatal Encounters*, a crowdsourced database of fatal use of force events. Chicago is one of the largest cities in the U.S. with relatively high rates of violent crime and a relatively large police force. Seattle has above average property crime rates and below average poverty rates. Both cities have higher rates of fatal police use of force than the typical large city.

We examine use of force by police officers occurring during arrests and, accordingly, the sample selection problem of interest is the concern that officers may differentially make arrests of potential arrestees across racial groups. Force events require in-person interactions between civilians and officers. When an officer arrests an individual, the event can involve resistance from the civilian, combativeness between the civilian and officer, and/or aggression on the part of an officer, each of which can contribute to a force event. Given these dynamics, researchers have often examined force outcomes coinciding with arrests (Fryer Jr, 2019; Weisburst, 2019) and we mirror this focus in our analysis. A large share of use of force incidents occur during arrests; in our data, 62 percent and 80 percent of force events can be linked to an arrest in Chicago and Seattle, respectively.[8]

[8]For the set of force incidents that are not linked to an arrest, only a small share have an

Another motivation for our focus on force during arrests is that analyzing disparities at the arrest margin is interesting in its own right. Although disparities at this stage have received substantially less attention than disparities at other stages of the criminal justice system in the literature – in part because of the empirical challenge of identifying a comparison group of individuals who are at risk of an arrest – figure 1 highlights that this margin may be especially important for understanding racial disparities in the criminal justice system more broadly. In the U.S., Black civilians comprise 12 percent of the overall population but about 33 percent of the incarcerated population. Black civilians also comprise about 12 percent of police-initiated contacts in the Police-Public Contact Survey but make up 26 percent of all arrestees according to the FBI. In other words, a sizable share of the well-known racial disparities observed at later stages of the criminal justice system appears to emerge at the arrest stage.

3.2 Data sources

For both Seattle and Chicago, our data include administrative records of each arrest made, each incident of police use of force, and records of each shift an officer works, typically called the "watch" data. Our data from Chicago cover the years 2012–2015 and our data from Seattle cover the years 2019–2022.

The arrest data include information on the charge or offense type, and both the arrest and force data include time stamps and information on location, involved officer(s), and some basic demographics on the arrestee or force victim. The watch data include each shift an officer works, including their rank during that shift (i.e., police officer or sergeant), shift start and end times, and assigned geography. These data are crucial for our empirical approach because they allow us to observe when officers work but do not make arrests, permitting the measurement of variation across officers in the propensity to make arrests. We link each arrest and each force incident to an officer × shift in the watch data using information on the officer(s) involved and the date/time of the incident. We also link arrests to force incidents using a combination of incident numbers, date/time, and involved officers.

Although Chicago and Seattle use slightly different names for the various notions of police geography, we adopt a harmonized terminology throughout the paper for consistency. We use *division* to refer to the largest sub-city areas ("police area" in Chicago and "precinct" in Seattle), *sector* to refer to the largest areas within divisions ("district" in Chicago and "sec-

identifiable originating event. In Chicago and Seattle, just 1 and 5 percent of force incidents originate with a non-arrest incident, respectively.

tor" in Seattle), and *beat* to refer to smallest geographic division ("beat" in both settings).[9] Append figure A-3 illustrates the relevant police geographies in both cities.

For our empirical analysis, we construct two primary datasets. The first is a panel dataset at the officer × shift level (i.e., one observation for each shift an officer works). This dataset includes information on the shift's time of day, day of week, assigned sector, and number of arrests made during each shift. We use only shifts worked at the rank of "police officer" (i.e., dropping detectives or commanding officers) and require that an officer work 100 such shifts for inclusion in our analysis sample.

The second is a dataset at the arrestee level (i.e., one observation per arrest). This dataset includes information on the date, time, and location of the arrest, information on the age, race, and gender of the arrestee, information on the criminal offense, the identity of the officer(s) making the arrest, and whether force was used during the arrest. We keep only arrests which can be linked to an officer × shift included in the above panel dataset.

For our baseline analysis, we pool the data from Chicago and Seattle together (although all our fixed effects will be allowed to vary across settings). Where relevant, we test the validity of stacking the data together and also show our core results separately by setting or using alternative approaches relying only on within-city analytical approaches. Our pooled officer × shift panel dataset includes approximately 2.2 million shifts worked by just under 5,000 unique officers. Our pooled arrestee dataset includes 134,361 arrests, which covers about 65 percent of the total arrests in our data (the majority of excluded arrests are those made by detectives; the remainder are those made by officers who do not work a sufficient number of shifts for inclusion in our panel dataset). Appendix table A-2 reports summary statistics for this analysis sample of arrestees.

3.3 Disparities in the selected sample

Table 1 examines racial disparities in police use of force in the selected sample of arrestees. Specifically, we report results from a series of regressions where the outcome of interest is whether an arrest results in police use of force. As shown in column (1), Black arrestees are 0.6 percentage points more likely to face force, or about 25 percent more likely relative to a non-Black average force rate of 0.023.

Controlling for detailed fixed effects at the level of the beat × $\mathbf{1}$[weekend] × time of day and division × year × month (which are the fixed effects we use in our baseline empirical

[9] In Chicago, there is an additional geography between district and beat. We use districts for the mid-tier geography ("sector") (i) because these align well with the geographic scope of sector in Seattle and (ii) district is comprehensively identified in the Chicago datasets.

specification, as described below) reduces the disparity slightly to to 0.55 percentage points. Adding in other demographics (age and gender) as controls slightly increases the estimated disparity, while further controlling for crime type reduces the gap to 0.47 percentage points.

The disparity reported in column 2, which conditions on our "design" fixed effects but not on other non-race characteristics, is the disparity we focus on adjusting for differential selection into the data. Figure 2 visually depicts the idea of our selection adjustment procedure, represented by equation (4). In this figure, the horizontal axis denotes the Black population share and the vertical axis captures the hypothetical force disparity between Black and non-Black individuals in the target sample. The solid blue circle corresponds to the moments from our selected sample of arrestees: 56 percent of arrestees are Black and Black arrestees are 0.55 percentage points more likely to face force. The solid blue line reports $\tilde{\Delta}$, the selection-corrected disparity, as a function of π, the Black share of the target sample. As π moves below the fraction Black in the selected sample π_s, indicating a higher selection probability for Black potential arrestees, we adjust the disparity upward (and vice versa). Our central empirical goal is to estimate π, the Black share of the population at risk of arrest.

4 Empirical approach

4.1 Measuring officer enforcement intensity

The first step in our empirical analysis is to measure variation in the propensity to make arrests, which we term enforcement intensity, across officers. To do so, we use our panel dataset at the officer $\times$ shift level and estimate the following regression:

$$N_{jt} = \alpha_j + \phi_s + \kappa_{dt} + u_{jt} \tag{5}$$

where N_{jt} is the number of arrests made by officer j during a given shift t,[10] the α_j's are officer fixed effects, the ϕ_s's are assignment fixed effects at the level of the assigned sector $\times$ shift (time of day) $\times$ weekend versus weekday, and the κ_{dt}'s are division $\times$ year $\times$ month fixed

[10]To avoid double-counting arrests in this first-stage regression, we divide each arrest by the number of arresting officers when computing N (e.g., if an officer makes one arrest in a given shift which is shared with one other officer, $N = 0.5$). We show that results are unchanged when using other measures of N in appendix table A-4.

effects.[11] From this regression, we compute each officer's adjusted enforcement intensity:

$$\tilde{N}_j = \hat{\alpha}_j + E[\hat{\phi}_s + \hat{\kappa}_{dt}],$$

or an officer's expected number of arrests per shift when working in the average assignment.

In panel (a) of appendix figure A-4, we report the distribution of estimated $\tilde{N}_j$'s across officers. The average officer makes 0.6 arrests per shift, but the distribution is very right-skewed, with an officer at the 99th (100th) percentile making 0.28 (0.74) arrests per shift. Figure A-4 also provides evidence that this between-officer variation represents "true" variation (as opposed to estimation error) by illustrating that an empirical Bayes shrinkage procedure (Morris, 1983) has only minimal impacts on the degree of dispersion across officers. Consistent with Ba et al. (2021) and Weisburst (2024), appendix table A-3 shows that, based on this measure of adjusted arrest volume, white, male, and younger officers tend to exhibit more intense enforcement behavior.

Panels (b) and (c) of appendix figure A-4 provide further validation of these officer enforcement intensity estimates. In panel (b), we estimate $\tilde{N}_j$'s for separate partitions of the data based on patrol locations and time and show that an officer's estimated arrest activity in one location or time period is highly predictive of their arrest activity in other locations and time periods. In panel (c), we provide a "first-stage" style estimate by randomly partitioning each officer's shifts into two groups, estimating a $\tilde{N}_j$ in each partition, and then regressing an officer's arrest activity on their estimated $\tilde{N}_j$ in the opposite partition, controlling for assignment and division-time fixed effects.

Our goal is to assign each arrest the enforcement intensity of the arresting officer. A minor complication arises from the fact that many arrests in our data involve multiple officers (as depicted in appendix figure A-2). To construct an arrest-level enforcement intensity measure, we follow the approach of Amaral et al. (2023) and take the average of the estimated $\tilde{N}_j$'s among arresting officers, weighting by each officer's total number of shifts. This feature of arrests means that the enforcement measure actually varies at the level of a *group* of officers, $g \in \mathcal{P}(\mathcal{J})$, where $\mathcal{P}(\mathcal{J})$ represents all subsets of $\mathcal{J}$. For simplicity of notation and exposition, we continue to denote enforcement intensity as $\tilde{N}_j$ and refer to "officer-level" enforcement, with the understanding that the variation is actually at the officer-group level.[12]

Figure 3 provides further validation for our approach by illustrating the relationship be-

[11] As discussed in section 3.2, sectors are sub-units of divisions in both cities. In Chicago (Seattle), there are 5 (5) divisions and 22 (17) sectors. See appendix figure A-3 for the relevant maps.

[12] As robustness, we consider the sensitivity of our main estimates to alternative approaches for aggregating the officer-level $\tilde{N}_j$'s into an arrest-level measure, including approaches which abstract from the need for aggregation by assigning one officer to each arrest, in appendix table A-4.

tween this arrest-level enforcement intensity measure and the crime type associated with each arrest. The likelihood that an arrest is for a serious crime declines sharply with our enforcement intensity measure, suggesting that high-enforcement officers are making additional arrests for less serious crimes relative to low-enforcement officers.

4.2 Identification of π

Our strategy for identifying π, the Black share of the target population, relies on this variation in enforcement intensity across officers and builds on the "identification at infinity" ideas advanced in recent research (e.g., Hull 2020, Arnold et al. 2022). The spirit of our approach is to use the racial composition of the most enforcement-intensive officers as our estimate of π. We discuss estimation in further detail in section 4.3 below.

Our approach relies on two assumptions. First, we require an exogeneity assumption which states that, within assignments (which we denote by the shift-time pair s, t), officers are randomly matched to potential arrests. Note that this *does not* imply that officers are randomly assigned to *arrestees*. Instead, this assumption is equivalent to stating that officers working in the same assignment encounter the same set of *potential arrestees*. And second, we assume the existence of some "extremum" officer j^* such that, if any other officer $j \neq j^*$ arrests an individual, the extremum officer j^* also arrests that individual. Stated formally, these assumptions are:

1. *Exogeneity*: $(R_i, X_i, \theta_i, \{S_{ij}, D_{ij}\}_{j\in\mathcal{J}}) \perp j_i \ \mid \ (s,t)$.
2. *Extremum-Agent Monotonicity*: $\exists\ j^* \in \mathcal{J}$ s.t., $\forall i, j,\ S_{ij} = 1 \Rightarrow S_{ij^*} = 1$.

Now, consider the set of individuals who hypothetically would be selected by the extremum officer j^*. These individuals are weakly a subset of the target population as defined in section 2.3, since $S_{ij^*} \leq \max_{j\in\mathcal{J}} S_{ij}$. By extremum-agent monotonicity, all individuals selected by any officer would be selected by j^*, so we also have that $\max_{j\in\mathcal{J}} S_{ij} \leq S_{ij^*}$. Therefore, the extremum officer's pool of (hypothetically) selected individuals *is* the target sample of all potential arrestees, $S_{ij^*} = \max_{j\in\mathcal{J}} S_{ij} = P_i$. Further, because of random assignment of officer-civilian encounters, individuals who are observed as arrested by officer j^* are in expectation identical to the target sample. Their characteristics can thus be used to identify the racial composition of the target sample:

$$Pr[R_i = b \mid j_i = j^*, S_i = 1] = Pr[R_i = b \mid S_{ij^*} = 1] = Pr[R_i = b \mid P_i = 1],$$

where the first equality follows from exogeneity, and the second equality follows from extremum-agent monotonicity.

As described further in section 5.3 and appendix B, the joint assumptions of exogeneity and extremum-agent monotonicity yield testable implications about the relationship between the composition of any given officer's selected sample of arrestees and the composition of the extremum officer's selected sample. We implement an associated test of these implications and cannot reject the joint null of exogeneity and extremum-agent monotonicity.[13]

Note that, while our use of variation in officer enforcement behavior for identification is similar in spirit to the examiner instrumental variables design (e.g. Kling, 2006; Chyn et al., 2024), our approach requires weaker assumptions, paralleling the discussion in section 2.4. In particular, our montonicity assumption is weaker that the strict monotonicity of Imbens and Angrist (1994).[14] Moreover, we do not impose an exclusion assumption requiring that the encountered officer only affects force through the arrest decision. As we discuss in appendix B.5, the lack of an exclusion restriction is a feature of our estimand of interest, which differs from the typical estimand in studies correcting for sample selection bias.

4.3 Estimation and inference

In practice, each officer's individual sample of arrestees is too small to offer sufficient statistical precision. To estimate the racial composition of the officers with the greatest enforcement intensity, we will exploit information from all officers, estimate the relationship between officers' racial composition and their enforcement intensity, and estimate a fitted value for the racial composition at the highest observed enforcement rate. Specifically, letting $Pr[R_i = b|S_i = 1, \tilde{N}_j]$ denote the share Black of arrestees among officers with average enforcement activity $\tilde{N}_j$, our estimate of π will be the value of this relationship at $\tilde{N}_{j^*} = \max_{j \in \mathcal{J}} \tilde{N}_j$:

$$\pi = \lim_{\tilde{N}_j \to \tilde{N}_{j^*}} Pr[R_i = b \mid S_i = 1, \tilde{N}_j]$$

Intuitively, we take the (selected) data on arrests and examine how the likelihood that an arrestee is Black, $Pr[R_i = b|S_i = 1, j_i = j]$, varies with the likelihood they were selected into the data based on the enforcement intensity of the arresting officer(s). We then rely on that

[13] Appendix figure A-5, which depicts the motononic relationship between an officer's overall and crime-specific enforcement intensities, provides additional evidence in support of our extremum-agent monotonicity assumption. Consistent with figure A-5, appendix table A-10 illustrates that our estimate of π is unchanged when estimating within-crime type and using officers' crime-specific arrest propensities.

[14] Specifically, strict monotonicity implies extremum-agent monotonicity but not vice versa. As an illustrative example that the converse is not true, extremum-agent monotonicity allows officers with the same selection propensity to have different arrestee demographic characteristics, and this is ruled out by strict monotonicity (Frandsen et al., 2023).

variation to estimate $Pr[R_i = b|S_i = 1, j_i = j]$ when $\tilde{N}_j = \tilde{N}_{j^*}$. Per the discussion in section 4.2, the Black share of arrestees for officers with the highest probability of arrest identifies the Black share of the target population.

We operationalize this idea with regressions of the form:

$$\mathbf{1}[R_i = b] = f(\tilde{N}_j) + \phi_s + \kappa_{dt} + \epsilon_i \tag{6}$$

where $\tilde{N}_j$ is the enforcement intensity constructed by aggregating the officer-level estimates (as described in section 4.1), and ϕ_s and κ_t are assignment and division-time fixed effects as above.[15] Note that this regression formulation, which adjusts for assignment by linearly including the fixed effects as controls, imposes an additional auxiliary linearity assumption, as discussed in Arnold et al. (2022). We probe the robustness of our results to relaxing this assumption below in section 5.2.

As our baseline approach, we specify a very flexible functional form for (6). Specifically, we adopt the semiparametric conditional binscatter approach of Cattaneo et al. (2024) and compute the fixed effects-adjusted average Black share of arrestees for 100 quantile bins of $\tilde{N}_j$, using the estimate from the top percentile as our estimate of π. Results from specifications specifying a linear functional form yield similar results, and we report estimates varying the quantile which we deem to be "extreme" for both binscatter and linear approaches.

To avoid potential empirical issues arising from the reflection problem or correlated errors in our first and second stages, we use a cross-partition approach when estimating π. Specifically, we randomly subset each officer's shifts into two partitions and estimate an enforcement intensity for each officer $\times$ partition using equation (5). We then use the enforcement intensity estimates from the opposite partition to construct the arrest-level enforcement intensity measure $\tilde{N}_j$ used in our second stage estimation.

With an estimate of π in-hand, it is straightforward to compute the selection-adjusted disparity in force $\tilde{\Delta}$ using equation (4). For statistical inference, we use a Bayesian boostrap (Rubin, 1981), clustering at the assignment (sector $\times$ day of week $\times$ shift) level. The randomized partitions used to construct our arrest-level enforcement intensity measure in the second stage are redrawn in each bootstrap replication.

[15] In our baseline approach, we replace sectors with the narrower notion of beats in the assignment fixed effects for this "second stage" regression. This choice provides modest precision gains but has no impact on point estimates, as shown in appendix table A-4.

5 Results

5.1 Baseline results

Figure 4 depicts the Black share of the selected sample (arrestees) as a function of enforcement intensity, adjusted for assignment and division-time fixed effects. This figure corresponds exactly to the binscatter version of our second stage regression (6) using 100 quantile bins. The horizontal dashed line indicates the Black share in the selected sample, $\pi_s = 0.5622$.

We find a strikingly downward-sloping relationship between $Pr[R_i = b|S_i = 1]$ and enforcement intensity, with a linear slope of $\beta = -0.165$ (0.028). The vertical height of the maximimal bin – which represents the strata-adjusted $Pr[R_i = b|S_i = 1]$ for officers in the top one percent of the enforcement intensity distribution – corresponds to our baseline estimate of $\pi = 0.494$. In other words, we find that while Black civilians comprise 56 percent of the selected sample arrestees, they comprise only about 49 percent of the target sample of potential arrestees. The estimated difference between π_s and π is statistically significant at conventional levels.

In table 2, we report the selection-adjusted disparity in use of force ($\tilde{\Delta}$) implied by this estimated over-selection of Black potential arrestees. In the selected sample, the disparity in use of force is $\Delta_s = 0.0055$ (0.001), while our estimated selection-adjusted disparity $\tilde{\Delta} = 0.011$ (0.002). In other words, the degree of differential selection into the sample shown in figure 4 implies a selection-adjusted force disparity which is *twice as large* as the disparity in the selected sample.

Table 2 also reports results using a linear specification for estimating π, obtained by computing the predicted value of the fitted line presented in figure 4 at the midpoint of the maximal enforcement intensity bin. Unsurprisingly given the patterns in figure 4, these results are nearly identical to those obtained using the bincsatter approach.

As discussed in section 2, we can compute a disparity in the likelihood of being arrested among potential arrestees, $Pr[S_i = 1|R_i = b] - Pr[S_i = 1|R_i = W]$, and the unscaled, unselected force disparity Δ with an estimate of one additional parameter, the overall probability of selection $Pr[S_i = 1]$. Following from our assumption that the most enforcement-intensive officers select potential arrestees with probability one, we obtain an estimate of the overall probability of selection by dividing each officer's estimated $\tilde{N}_j$ by the maximum $\tilde{N}_j$ and then taking the shift-weighted mean of this rescaled $\tilde{p}_j$.

Appendix table 3 reports the implied disparities in the target sample. Specifically, our estimates indicate that Black potential arrestees are about 3 percentage points more likely to be arrested, relative to a non-Black arrest rate of 9.6 percent. Among potential Black and non-Black arrestees, the probability of force is 0.0036 and 0.0022, respectively. This difference is statistically significant at conventional levels and implies that a Black potential arrestee faces a 64 percent higher risk of force than a non-Black potential arrestee.

5.2 Robustness

In appendix table A-4, we explore the sensitivity of our baseline binscatter estimates to our specification choices. In panel (a), we report results when using alternative measures of enforcement intensity in the first stage regression (5). In panel (b), we report results using alternative methods for aggregating the officer-level enforcement intensity measures into an arrest-level measure, including taking the minimum and maximum among the arresting officers and randomly drawing one officer when there are multiple arresting officers. In panel (c), we consider alternative fixed effects, including using beats instead of sectors in the first stage where applicable. This table illustrates clearly that none of these choices have any bearing on our empirical conclusions.

We also consider the importance of estimation error in officer enforcement intensity for our findings. Note that estimation error plays two conflicting roles in our empirical approach. On one hand, measurement error in enforcement intensity will tend to attenuate relationship depicted in figure 4, leading us to overestimate π (understating the extent of selection bias). On the other hand, estimation error generates an over-dispersed distribution of enforcement intensity, potentially leading us to overestimate extreme quantiles in this distribution and thereby underestimate π (overstating the extent of selection bias).

To assess the importance of estimation error for our conclusions, we adopt the following two-step procedure. First, we use a split-sample IV (SSIV) approach, where we estimate officer enforcement intensity in two random partitions of the data and instrument an officer's estimate in one partition with their estimate from the opposite partition, to obtain an estimate of the linear slope in the relationship between $Pr[R_i = b|S_i = 1]$ and enforcement intensity (i.e., the relationship depicted in figure 4) which is corrected for attenuation bias.

We then construct bounds on π by extrapolating this unattenuated slope estimate to various quantiles in the distribution of enforcement intensity. Appendix table A-5 reports the SSIV estimates and table A-6 reports the associated bounds.[16] Our most conservative bound, which extrapolates to the 99th percentile of the empirical Bayes shrunken distribution

[16] As reported in appendix table A-5, the measurement error corrected slope we obtain using the SSIV approach is only about five percent larger (more negative) than the slope we report in figure 4, suggesting that enforcement intensity is quite well-measured.

of enforcement intensity (which should be strictly less than the true maximum enforcement intensity), yields $\pi = 0.537$ and an associated $\tilde{\Delta} = 0.0077$. This estimate is comparable to the lower 95 percent confidence bound of our baseline estimate and is still about 40 percent larger than the selected disparity. On the other hand, using the unattenuated slope but extrapolating to extreme quantiles in the unshrunken distribution of enforcement intensity (which should be strictly greater than the true maximum) yields $\pi = 0.41$ and an associated $\tilde{\Delta} = 0.016$, a disparity about three times larger than that in the selected sample.

Along similar lines, figure 5 explores how our conclusions would change when using alternative quantiles to define the maximally-enforcing officers. Panel (a) reports estimated π's with 95 percent confidence bands for both binscatter and linear specifications when using the top q percent of the enforcement distribution to estimate π, where q is on the horizontal axis. Relative to our baseline estimate ($\pi = 0.4929$), estimated π's are generally slightly larger (i.e., closer to $\pi_s = 0.562$) when using less extreme quantiles and and slightly smaller (i.e, further from π_s) when using higher quantiles. Note that this pattern is expected given the linear relationship depicted in figure 4. Panel (a) of figure 5 also reports the estimated π when using data-driven optimal bin selection of Cattaneo et al. (2024) to select the bins ($\pi = 0.5048$), which is very similar to our baseline estimate ($\pi = 0.4938$).

The second panel of figure 5 reports the corresponding selection-adjusted force disparities $\tilde{\Delta}$ when using the quantile-specific estimated π. By construction, the pattern in panel (b) is inverse of the pattern in panel (a); as the estimated π becomes further from π_s, the selection-adjusted $\tilde{\Delta}$ grows relative to the selected sample Δ_s. When using the most conservative π from the first panel, we obtain an adjusted force disparity $\tilde{\Delta} = 0.0076$, about forty percent larger than the selected sample disparity ($\Delta_s = 0.0055$). When using the least conservative estimate ($\pi = 0.451$) which linearly extrapolates to the 99.99th percentile in the enforcement intensity distribution, the estimated adjusted force disparity is $\tilde{\Delta} = 0.0139$, about 2.5 times as large as the selected disparity. Note that, in all cases, the estimated difference $\tilde{\Delta} - \Delta_s$ is statistically significant at conventional levels.

Another concern for our approach is the worry that officers in different parts of the enforcement intensity distribution are simply patrolling different areas, which would violate an auxiliary linearity assumption implicit in our second stage estimation (Arnold et al., 2022). To address this concern, we estimate π using an alternative, within-locations approach based on the procedures in Feigenberg and Miller (2022) and Goncalves and Mello (2023). Specifically, for varying notions of location, we first estimate officer-by-location enforcement intensity. Then, using these location-specific enforcement intensity measures, we conduct location-specific extrapolations (i.e., estimate the second stage separately by location) and construct estimates of π relying only on this within-location variation by aggregating up the location-specific estimates, weighting by location shares in the selected sample.

We depict the results of this approach and report the within-location estimated π's in figure 6. The slope in the relationship between $Pr[R_i = b|S_i = 1]$ and enforcement intensity is strikingly similar when using only within-city, within-division, and within-sector variation. Accordingly, we obtain very similar estimates of π using these alternative approaches. Accompanying figure 6, appendix figure A-6 shows the city-specific versions of figure 4. While Chicago and Seattle differ meaningfully both in terms of the Black share of arrestees and the distribution of enforcement intensity, we find a strikingly similar relationship between enforcement intensity and the race composition of arrestees in both settings. We cannot statistically reject that the slope in this relationship is equal in the two cities.

An alternative concern we note is that some arrests involving use of force may initiate with civilian choices to confront officers rather than officer choices to arrest civilians. In appendix table A-7, we report estimates when dropping arrests for officer assault from our analysis sample, which are nearly identical to our baseline estimates.

Finally, we acknowledge worries about the measurement of police use of force, which are pervasive in the literature on this topic. Recall that measurement of use of force is not an input into our primary estimation exercise. Rather, our approach takes the race-specific average force rates in the selected sample and adjusts the disparity therein for the estimated degree of differential selection into the data by race. Hence, our approach can easily be adapted to report adjusted disparities under various assumptions about the mismeasurement of police use of force. To speak more directly to this question, we report estimates for only use of force which results in civilian injury, more severe force events that we expect are less subject to misreporting concerns, in appendix table A-8. In the selected sample, non-Black civilians are injured by the police in 0.58 percent of arrests and Black arrestees are 0.07 percentage points more likely to be injured. The selection-corrected disparity is 0.21 percentage points, or about three times larger than the selected disparity.

5.3 Testing assumptions

As introduced in section 4.2, our approach for identifying the race composition of potential arrestees using an extremum officer j^* requires a monotonicity assumption stating that, for all officers j, $S_{ij} = 1$ implies $S_{ij^*} = 1$. Another way to phrase this assumption is that the set of arrestees for each officer is a subset of the set of arrestees for the extremum officer.

Combined with our assumption of exogeneity, this notion of monotonicity implies testable bounds on the composition of arrestees for each officer in the data. The underlying logic of these bounds is that each officer's arrest rate imposes limits on how much her sample composition can differ from that of the extremum officer. For example, if extremum-agent monotonicity holds, an officer who arrests 99 percent as often as the extremum officer must have a share Black among her arrestees which lies between $\pi \times 0.99$ and $\pi \times 0.99 + 0.01$, where the lower (upper) bound assumes that the "missing" arrestees are all Black (non-Black) civilians. Formally, defining π_j to be the fixed-effects adjusted estimate of officer j's Black share of arrestees and $\tilde{N}_j$ to be their enforcement intensity, the joint assumptions of exogeneity and extremum-agent monotonicity imply:

$$\pi_j \tilde{N}_j \leq \pi \tilde{N}_{j^*} \leq \pi_j \tilde{N}_j + (\tilde{N}_{j^*} - \tilde{N}_j),\ \forall j \in \mathcal{J}$$

Hence, we can test the combined assumptions of exogeneity and extremum-agent monotonicity by jointly testing whether these inequalities hold for all officers in the sample. We implement this as a joint test of moment inequalities using the inference procedure developed in Romano et al. (2014) and Bai et al. (2022), as described further in appendix B. Table A-9 reports the results of this test separately for each arrestee characteristic as well as the results of a joint test for all arrestee observables. We cannot reject the null hypothesis that the above inequalities hold when focusing only on racial composition ($p = 0.24$) or when jointly examining all arrestee characteristics ($p = 0.12$), suggesting that our assumptions are likely to hold in our setting.[17]

As an additional test of the validity of our extremum-agent monotonicity assumption, we recompute π by estimating officers' race-specific enforcement intensity $\tilde{N}_j^r$, taking the exremum officers' values $\tilde{N}_{j^*}^r$, and computing $\hat{\pi} = \tilde{N}_{j^*}^b / (\tilde{N}_{j^*}^b + \tilde{N}_{j^*}^w) = 0.512$. The similarity of this estimate with our baseline estimate suggests that the extremum officer captures the most extreme officers in terms of arrests of both racial groups.

5.4 Alternative populations of interest

Our analysis focuses on the population at risk of arrest as the primary target population of interest. However, we also note that not all use of force events occur during an arrest and that racial disparities in the risk of facing police use of force among alternative target populations may similary be of interest.

[17] In appendix B, we also discuss alternatives to our monotonicity assumption. In particular, without extremum-officer monotonicity, we propose an alternative ϵ-monotonicity assumption which states intuitively that a hypothetical officer satisfying extremum-officer monotonicity would need to select at a rate ϵ higher than the most enforcing officer in the data. This ϵ-monotonicity assumption suggests extrapolation (e.g., Arnold et al. 2022) or bounding approaches for estimating the composition of the unselected sample of interest.

It is straightforward to compute our selection-adjusted $\tilde{\Delta}$ using population benchmarks. Based on data from the 2010 Census, the combined population of Chicago and Seattle is 28 percent Black (as reported in figure 2). For this notion of the target population (i.e., letting $\pi = 0.28$), we estimate $\tilde{\Delta} = 0.022$, about twice as large as our estimate based on the sample of potential arrestees and four times lager than the disparity in the selected sample. We can also compute the "beat-weighted" share Black of city residents, where we reweight city neighborhoods to match their relative frequencies in the arrest data. As shown in figure 2, this gives $\pi = 0.428$ and implies $\tilde{\Delta} = 0.015$, about 35 percent larger than our baseline estimate using potential arrestees.

An alternative target sample of interest is set of individuals who are at risk of any police contact, rather than those at risk of an arrest, which in principle would represent the full population of individuals at risk of force. If our data included all police-civilian encounters, we could replicate our econometric procedure leveraging variation across officers in the propensity for any civilian contact, instead of the propensity to make arrests, to recover the racial composition of this alternative sample of interest. In practice, however, the universe of police-civilian interactions is never captured in administrative datasets.

There are two ways we could speak to this question. First, we note that the most comprehensive national dataset on police interactions, the Police-Public Contact Survey, suggests that racial composition of individuals with any police contact or any police-initiated contact is quite similar to the racial composition of the overall population, as depicted in figure 1. Hence, population benchmarks may be reasonably informative about the composition of those at risk of police contact, suggesting that the unselected disparity for potential arrestees that we recover could be a conservative estimate of the disparity among those with the potential for any police contact.

Alternatively, we could note that arrests are a form of police-civilian contact and, therefore, that the set of potential arrestees is a strict subset of the population at risk of any police contact. Hence, one could construct bounds on the racial composition of the sample at risk of contact by taking our estimated π and then making assumptions about the fraction of interactions resulting in arrest and the race composition of the "missing" interactions. Simple "best" and "worst" case bounds are likely to be uninformative because arrests comprise a small subset of all police-civilian interactions, but bounds could potentially be tightened by drawing on additional data sources or by making additional assumptions. For example, it may be reasonable to assume that the over-selection of Black civilians is weakly more pronounced for lower-level level encounters, which are subject to less downstream oversight, than for arrests. Such an analysis is beyond the scope of our paper but represents a potentially interesting avenue for future research.

A potential critique of our notion of the target sample might be that the most-enforcing officers are "over-arresting," or making many illegitimate arrests which should not be included in the target sample. This critique may stem from a researcher's interest in defining the target sample in more restrictive terms, e.g., among a sample of externally-validated potential arrests, in place of our focus on the full set of potential arrest interactions. Such a concern would imply that the target sample P_i is "too large," and we can place bounds on the racial composition of any subset of the target sample we identify. Letting π^{ρ} denote the Black share of a subsample of P_i, where ρ is the probability of being in this subsample conditional on being in P_i, we know that $\pi^{\rho} \in [\frac{\pi}{\rho} - \frac{1-\rho}{\rho},\ \frac{\pi}{\rho}]$. For example, if $\rho = 0.9$, so that the desired sample is 10 percent smaller than the set of potential arrestees, the bounds are $[0.438, 0.549]$. Since this interval is strictly below the Black share in the selected sample of arrestees, our central conclusion that the naive disparity in the selected sample is biased downwards would not change for any target sample definition that is at least 90 percent as large as our target sample of all potential arrestees.

A strength of our target sample is that it corresponds to those actually at risk of force, given the arresting behavior of the observed set of officers. However, also worth considering is how our target sample would compare to a normative notion of a population of interest. The existing literature has taken an implicitly normative stance on the choice of target population; as the problem is stated in Knox et al. (2020), "if police racially discriminate when choosing whom to investigate, analyses using administrative records to estimate discrimination in police behavior are statistically biased." Given the word *discrimination*, the implication here is that the "correct" target population is one where, after accounting for other demographic and incident characteristics, race is not a factor in determining entry into the target population.

We provide some evidence indicating that our target population of potential arrestees satisfies this normative criterion in appendix table A-12. Specifically, this table illustrates the racial differences in the characteristics of individuals in the selected sample of arrestees ($P_i = 1, S_i = 1$), individuals in the target sample of potential arrestees ($P_i = 1$), and individuals who are in the target sample but not arrested ($P_i = 1, S_i = 0$). As illustrated in the first panel, we find significant differences across race among those who are arrested. The second panel illustrates that these differences are dramatically less pronounced when examining the full target population, although we can still reject the null that all characteristics are equal across racial groups at the 10 percent level. Among potential arrestees who are not arrested, however, we fail to reject the null that non-race characteristics are equal across racial groups. The not-selected individuals ($P_i = 1, S_i = 0$) are those who are not arrested by their encountered officer but *would* be arrested by the most-enforcing officers. They can thus be seen as approximating the population who would only be arrested by an officer who is marginally *more-enforcing* than the current extremum officer, a population that is therefore "at the margin" of entering the target sample. Hence, the similarity of non-race characteristics across racial groups among this not-selected sample supports the view that, conditional on other characteristics, race is not a factor in determining entry into the target population.

6 Discussion

6.1 Disparities versus discrimination

As noted in section 2, our analysis delivers an unselected racial disparity which does not necessarily correspond to a standard notion of *causal* discrimination, since force outcomes may be due to non-race characteristics that are correlated with race. Indeed, a large literature across various settings has documented that race disparities may be partly explained by differences in non-race characteristics.[18] In other words, our estimated Δ does not correspond to a racial disparity in treatment which is conditional on all other characteristics.

To conceptualize the relationship between Δ and other notions of discrimination, suppose that all non-race characteristics considered by officers can be represented by a single variable θ_i which takes discrete values. The unconditional disparity can be expressed with the standard Kitagawa-Oaxaca-Blinder decomposition:

$$\Delta \;=\; \sum_{\theta} \left[D^{b,\theta} - D^{w,\theta}\right] Pr[\theta_i = \theta] + \sum_{\theta} \left[\theta^{b} - \theta^{w}\right] \tilde{D}^{\theta}, \tag{7}$$

where $D^{r,\theta} = E[D_i | R_i = r, \theta_i = \theta]$, $\theta^{r,\theta} = Pr[\theta_i = \theta | R_i = r]$, and $\tilde{D}^{\theta} = \pi D^{w,\theta} + (1 - \pi) D^{b,\theta}$. The first term captures racial differences in how individuals with the same other characteristics θ_i are treated, and it corresponds closely to the idea of discrimination targeted in much of the literature. The second term captures racial differences in the composition of other characteristics and reflects the idea that differences in treatment may arise from, for example, differences in offense type.

The above decomposition also highlights that, if the race-specific distributions of θ are equal in the target sample (i.e. if, $\forall \theta,\ \theta^{b} = \theta^{w}$), then our estimated Δ corresponds to

[18]For a selective list, see Neal and Johnson (1996) on earnings, Fryer Jr (2011) and List and Uchida (2025) on academic achievement, and Jordan et al. (2024) on criminal justice involvement.

discrimination. While we do not directly observe θ_i, we can partially test this condition by examining racial differences in *observable* characteristics X among the target sample. Specifically, we can calculate the race-specific average characteristics using Bayes' Rule: $Pr[X = x|R = r] = Pr[R_i = r|X_i = x]Pr[X_i = x]/Pr[R_i = r]$. Our baseline estimation approach can identify $Pr[R_i = r]$ and $Pr[X_i = x]$. To identify the conditional race shares $Pr[R = r|X = x_i]$, we repeat our approach within discrete values of X.

Table 4 explores covariate differences across racial groups in both the selected and target samples. For simplicity, we report race-specific averages of a covariate index capturing an individual's predicted likelihood of facing force.[19] In the selected sample, Black arrestees are about 0.1 percentage points (four percent) more likely to face force based on their non-race covariates, with this disparity statistically significant at conventional levels. However, in the target sample of potential arrestees, this gap is an order of magnitude smaller and no longer statistically distinguishable from zero. In other words, we find evidence of meaningful differences in non-race characteristics among arrestees, but not among potential arrestees.

Table 4 also reports results from joint tests of the null hypothesis that covariates are equal are equal across racial groups in the selected and target samples. Aligning with the conclusions when examining the covariate index, p-values from these tests are consistently below 0.001 in the selected sample, suggesting statistically relevant differences in covariates across racial groups. In the target sample, on the other hand, we cannot reject that the distribution of offense types is equal across racial groups ($p = 0.38$) and can only reject the null that all covariates are equal at the ten percent level ($p = 0.06$).

To provide a more direct assessment of the relative importance of discrimination versus differences in non-race characteristics in driving our estimated disparity, we perform our selection-correction exercise *within* observable characteristics, which we describe in detail in appendix B. We construct a discrete variable X_i that divides the sample into sixteen cells based on arrestee gender, whether they are younger than 35, and four offense categories (violent, property, drug, other). Within each value of X_i, we perform our baseline estimation approach to identify $Pr[R_i = b|X_i = x]$ and use it to estimate $\Delta_x = D^{b,x} - D^{w,x}$. We then decompose the unconditional force gap into components that are within and across

[19] This covariate index is constructed by regressing an indicator for use of force on all non-race covariates using only non-Black individuals and then computing predicted values. As in the construction of our enforcement intensity instrument, we use randomized cross-partitions to construct predicted force (i.e., we estimate this regression separately in two partitions and construct predicted values using the regression coefficients from the opposite partition). Race-specific averages of all covariates for the selected and target samples are reported in appendix table A-12.

demographic cells:

$$\tilde{\Delta} = \underbrace{\sum_x \frac{\Delta_x}{E[S_i|R_i = b]} Pr[X_i = x]}_{\text{Within-}X\text{ racial force gap}} + \underbrace{\sum_x \frac{[X^b - X^w]}{E[S_i|R_i = b]} \tilde{D}_x}_{\substack{\text{Racial differences} \\ \text{in } X\text{-shares}}},$$

where $X^r = Pr[X = x|R_i = r]$ and $\tilde{D}_x = (1 - \pi)E[D_i|R_i = b, X_i = x] + \pi E[D_i|R_i = w, X_i = x]$. Averaging across demographic cells, and weighting by the size of the cells in the target sample, $Pr[X_i = x]$, we calculate an average within-X force disparity of 0.0074 (0.002), as shown in table B-1. This figure is 67 percent of our unconditional force disparity ($\tilde{\Delta}$), providing further evidence that the majority of the force gap cannot be explained by observable characteristics of the arrestee or the offense. This fact suggests an important role for officer discrimination in driving the racial force gap in the unselected sample.

It is well known that the KOB decomposition is not unique, and different decompositions correspond to different causal statements. As we discuss in appendix B, under the assumption of independence of race conditional on observables X, the above within-X force gap reflects the causal effect of race on the probability of force for all individuals in the target sample. An alternative estimand of interest is the causal effect of race on force for individuals in the *selected* sample of arrestees. We calculate the conditional force gaps for this group as well, and we again find similar magnitudes for the causal effect of race on force likelihood.

6.2 Revisiting Fryer Jr (2019)

A natural question following from our analysis is to what extent the differential selection that we document affects the highly-publicized conclusions in Fryer Jr (2019). This study examines racial differences in police use of force across a range of data sources and force outcomes and presents regression results which account for a rich set of controls for civilian and encounter characteristics. While Fryer Jr's analyses finds large racial disparities for less severe forms of force, he found no statistical difference in levels of severe or fatal force, and point estimates for these outcomes actually indicated greater force rates against white civilians.

As in most of the literature focused on policing, a limitation of data sources used in Fryer Jr (2019) is that they originate from incidents involving police discretion, such as pedestrian stops or arrests. This issue has attracted a large follow-up literature debating the possible importance of sample selection bias for the study's conclusions (Knox et al., 2020; Durlauf and Heckman, 2020; Fryer Jr, 2020). While we cannot directly replicate our procedure for estimating the race composition of the unselected sample using Fryer Jr's data, we can ask how the study's conclusions would change assuming a comparable rate of differential selection by race in his data and ours.

First, we note that the target estimand in Fryer Jr (2019) corresponds closely to the within-X racial disparity measure from the previous section. Combining the notation from sections 2 and 6.1, we can think of his estimand of interest as $\Delta^{\text{Fryer}} = \sum_x \omega_x [D^{b,x} - D^{w,x}]$, for some set of $\omega_x > 0$, where X represents a rich set of covariate controls. While this estimand represents within-X force disparities for the unselected sample, the data available are for a *selected sample of observations*, and thus Fryer Jr (2019) estimates a series of regressions that identify some form of

$$\Delta_s^{\text{Fryer}} = \sum_x \tilde{\omega}_x [D_s^{b,x} - D_s^{w,x}],$$

where $D_s^{r,x} = E[D_i | R_i = r, X_i = x, S_i = 1]$ and $\sum_x \tilde{\omega}_x = 1$. Following the logic of section 2 and applying a within-X version of equation (2), we can decompose the estimate from the selected sample to show how it differs from the target estimand:

$$\Delta_s^{\text{Fryer}} = \underbrace{\sum_x \tilde{\omega}_x \frac{\Delta_x}{S^{b,x}}}_{\substack{\text{Unselected} \\ \text{force disparity}}} + \underbrace{\sum_x \big[\frac{\pi^x}{1-\pi^x} \cdot \frac{1-\pi_s^x}{\pi_s^x} - 1\big] \tilde{\omega}_x D_s^{w,x}}_{\text{Sample selection bias}}$$

where $S^{r,x} = E[S_i | R_i = r, X_i = x]$, $\pi^x = Pr[R_i = b | X_i = x]$, and $\pi_s^x = Pr[R_i = b | X_i = x, S_i = 1]$. This decomposition shows that, if there is no differential selection by race conditional on X (i.e., $\pi^x = \pi_s^x$ for all x), Δ_s^{Fryer} captures a weighted average of within-X (scaled) force disparities, $\Delta_x / S^{b,x}$.

Fryer Jr (2019) reports various estimates of Δ_s^{Fryer} for different samples and force outcomes, as well as averages of race specific force. Naturally, the study does not report $D_s^{w,x}$ for each value of X. Therefore, we are not able to directly estimate the "sample selection bias" term above. However, we can resolve this issue by assuming that $\big[\frac{\pi^x}{1-\pi^x} \cdot \frac{1-\pi_s^x}{\pi_s^x} - 1\big]$ is the same across all values of x, denoting this with $\big[\overline{\frac{\pi^\cdot}{1-\pi^\cdot} \cdot \frac{1-\pi_s^\cdot}{\pi_s^\cdot}} - 1\big]$.[20] Making this simplification and rearranging terms, we arrive at an estimable expression for the average force disparity:

$$\sum_x \tilde{\omega}_x \frac{\Delta_x}{S^{b,x}} = \Delta_s^{\text{Fryer}} + \big[1 - \overline{\frac{\pi^\cdot}{1-\pi^\cdot} \cdot \frac{1-\pi_s^\cdot}{\pi_s^\cdot}}\big] \cdot \sum_x \tilde{\omega}_x D_s^{w,x}. \tag{8}$$

[20]To validate this assumption, we estimate this object separately for 16 groups at the level of gender × $\mathbf{1}$[age≥35] × crime type and test the null hypothesis that $\frac{\pi_x}{1-\pi_x} \cdot \frac{1-\pi_s^x}{\pi_s^x}$ is equal across these groups. We cannot reject that this object is equal across groups ($p = 0.69$).

The first term on the right-hand side represents Fryer's estimates of race on force in the selected sample. For the second term, we take $\sum_x \tilde{\omega}_x D_s^{w,x}$ from his reported summary statistics on use of force against white civilians in the selected sample and then construct the ratio of unselected and selected race shares directly from our analysis. Implicit in this calculation is the assumption of proportionally similar differential selection by race within observable characteristics in his setting and ours.

We report the results from this exercise in table 5. Each row presents a different data source and outcome examined by Fryer Jr (2019). The first column presents the study's reported average force outcome for white individuals in the (selected) sample. Columns (2) and (4) present the study's measures of racial force gaps without any controls and with the full set of available controls, respectively, and columns (3) and (5) present our selection-corrected estimates of the coefficients in columns (2) and (4). The first two rows in columns (2) and (4) show positive coefficients for the Black-white gap in force from the NYC Stop-Question-Frisk data and the Police Public Contact Survey (PPCS). Our correction increases the coefficients in all cases. Because average force rates are higher in the second row (column 1), selection adjustments are larger for the PPCS analysis; we find that the selection-corrected force gaps in the PPCS are about 40 percent larger than suggested by the selected sample.

The bottom three rows present coefficients from Fryer Jr's analysis of police shootings in Houston. Each row corresponds to a different choice of analysis sample. Note that all coefficients are negative, indicating that Black individuals in the sample are *less likely* to be shot by the police. While our selection correction increase all coefficients, in only one of six cases is the adjusted estimate positive (and, in this case, the magnitude is quite small). Accompanying table 5, appendix figure A-7 repeats our figure 2 but depicts the selected sample moments from the third row of table 5, corresponding to Fryer Jr's analysis of shooting in Houston which condition on features of the interaction drawn from police narratives. Correcting for sample selection bias shifts the estimate leftward along the selection adjustment curve sufficiently far such that the estimated disparity is approximately zero. The conclusion from this exercise is that, while our selection correction is sufficiently important to erode (to varying degrees) the presence of "reverse discrimination" observed among police shootings in Fryer Jr's analysis, it does not *reverse* his conclusions and result in meaningfully positive Black-white disparities for this margin of force.

6.3 Implications for other stages of the criminal justice system

While our analysis focuses on computing the racial disparity in police of force which is purged of selection bias arising from racial disparities in the likelihood of being selected into the data, worth noting is the point that these conclusions about differential selection into arrests have implications for our understanding of the observed racial disparities at various downstream stages of the criminal justice system.

As discussed in section 5.1, our findings suggest that Black potential arrestees are about 30 percent more likely to be arrested than non-Black potential arrestees. Our analysis in sections 6.1 and 6.2 show the similarity of non-race characteristics among Black and non-Black potential arrestees and suggest differential rates of selection within covariates which are similar to our baseline estimates. Collectively, these findings support the interpretation that differential arrest rates by race are in part determined by race discrimination by officers.

As an example to illustrate the implications of differential arrest rates by race for understanding disparities in other downstream criminal justice outcomes, consider the summary statistics reported in Dobbie et al. (2018)'s analysis of bail hearings. In their pooled sample based on data from Philadelphia and Miami, 58.3 percent of arrestees are Black and Black arrestees are about 5 percentage points more likely to be detained pretrial. Assuming the same proportional degree of differential selection into arrests in their setting and ours would imply a target sample $\pi = 0.512$ and a selection-corrected disparity in the likelihood of pretrial detention of $\tilde{\Delta} = 0.175$, or about three times larger than the disparity in the selected sample of arrestees.

Note that this exercise re-highlights two important features of our analysis: (i) our framework applies more generally to outcomes which are conditional on discretionary selection; (ii) our selection-corrected disparity can be calculated easily using information typically reported in summary statistics tables or popular press articles under various assumptions about the extent of differential selection into the data.

7 Conclusion

Estimating and understanding racial disparities in the criminal justice system is complicated by the fact that criminal justice agents have broad discretion at various stages of the system. Discretion exercised at one stage of the system influences which individuals appear in datasets covering downstream stages, potentially introducing sample selection bias which can distort measures of realized racial disparities.

In this paper, we study racial disparities in police use of force against civilian arrestees, taking into account the possibility of selection bias arising from racial disparities at the arrest stage. We first develop a conceptual framework to illustrate that a racial disparity which is purged of selection bias can be calculated from moments in the selected data and an estimate of the racial composition of the potentially-selected sample of interest.

We then propose an approach for estimating this racial composition which leverages variation across officers in their propensity to make arrests. Given an exogeneity assumption and a monotonicity-like assumption requiring the existence of some officer who makes all arrests that would be made by other officers, we show that the Black share of arrests made by the most-enforcing officers corresponds to the Black share of the population at risk of arrest, and therefore the population at risk at facing force during an arrest.

Implementing our approach using data on arrests and use of force from the Chicago and Seattle police departments, we find strong evidence of differential selection into the arrests data by race. While Black civilians comprise 56 percent of arrestees, we estimate that they comprise about 49 percent of potential arrestees. Accordingly, we find that after adjusting for selection bias, Black civilians are 48 percent more likely to face force than non-Black civilians, a disparity about twice as large as the naive difference computed in the selected data. We also find that about 70 percent of this disparity in within non-race observable characteristics, suggesting that racial discrimination by officers could play a meaningful role in driving the disparities that we document.

Our finding of significant racial disparities in arrest rates among potential arrestees has important implications for the interpretation and understanding of racial disparities in the criminal justice system more broadly. Given the unavailability of an observable control group, much of the prior literature has ignored the potential for differential selection at the earliest entry points in this system, such as arrests or stops made by police officers. Future work should continue to take seriously the ways in which initial choices by police officers can affect the population observed in data at later stages of the criminal justice system.

References

Adger, C., Ross, M., and Sloan, C. (2022). The effect of field training officers on police use of force. *United States Military Academy at West Point Working Paper*.

Aggarwal, P., Brandon, A., Goldszmidt, A., Holz, J., List, J. A., Muir, I., Sun, G., and Yu, T. (2025). High-frequency location data show that race affects citations and fines for speeding. *Science*, 387(6741):1397–1401.

Alpert, G., Smith, M., and Dunham, R. (2004). Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy*, 6(1).

Amaral, S., Dahl, G., Endl-Greyer, V., and Hener, T. (2023). Deterrence or backlash? arrests and the dynamics of domestic violence. *Unpublished manuscript*.

Ang, D. (2021). The effects of police violence on inner-city students. *Quarterly Journal of Economics*, 136(1):115–167.

Ang, D., Bencsik, P., Bruhn, J., and Derenoncourt, E. (2025). Community engagement with law enforcement after high-profile acts of police violence. *American Economic Review: Insights*, 7(1):124–42.

Antonovics, K. and Knight, B. (2009). A New Look at Racial Profiling: Evidence from the Boston Police Department. *Review of Economics and Statistics*, 91(1):163–177.

Anwar, S. and Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, 96(1):127–151.

Arnold, D., Dobbie, W., and Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038.

Ba, B., Knox, D., Mummolo, J., and Rivera, R. (2021). Diversity in policing: The role of officer race and gender in police-civilian interactions in Chicago. *Science*, 371(6530):696–702.

Bai, Y., Santos, A., and Shaikh, A. M. (2022). A two-step method for testing many moment inequalities. *Journal of Business & Economic Statistics*, 40(3):1070–1080.

Baron, J., Doyle, J., Emmanual, N., Hull, P., and Ryan, J. (2024). Discrimination in multi-phase systems: Evidence from child protection. *Quarterly Journal of Economics*, 139(3):1611–1664.

Bloomberg, M. (2013). Stop and frisk is not racial profiling. *Washington Post*.

Blundell, R., Gosling, A., Ichimura, H., and Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363.

Bohren, J., Hull, P., and Imas, A. (2025). Systemic discrimination: Theory and evidence. *Quarterly Journal of Economics*, 112(3):1743–1799.

Burghart, B. (2024). Fatal Encounters [Dataset].

Cai, W., Gaebler, J., Kaashoek, J., Pinals, L., Madden, S., and Goel, S. (2022). Measuring racial and ethnic disparities in traffic enforcement with large-scale telematics data. *PNAS Nexus*, 1(4).

Cattaneo, M., Crump, R., Farrell, M., and Feng, Y. (2024). On Binscatter. *American Economic Review*, 114(5):1488–1514.

Chalfin, A. and Goncalves, F. (2023). Professional motivations in the public sector: Evidence from police officers. *NBER Working Paper 31985*.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *journal of Econometrics*, 32(2):189–218.

Chyn, E., Frandsen, B., and Leslie, E. C. (2024). Examiner and judge designs in economics: A practitioner's guide. Technical report, National Bureau of Economic Research.

Desilver, D., Lipka, M., and Fahmy, D. (2020). 10 things we know about race and policing in the u.s. *Pew Research Center*.

Dobbie, W., Goldin, J., and Yang, C. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40.

Durlauf, S. N. and Heckman, J. J. (2020). An empirical analysis of racial differences in police use of force: A comment. *Journal of Political Economy*, 128(10):3998–4002.

Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., and Dijk, W. V. (2021). Selection in surveys: Using randomized incentives to detect and account for nonresponse bias. *NBER Working Paper Series*, 29549.

Feigenberg, B. and Miller, C. (2022). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *Quarterly Journal of Economics*, 137(1):49–113.

Frandsen, B., Lefgren, L., and Leslie, E. (2023). Judging judge fixed effects. *American Economic Review*, 113:253–77.

Fryer Jr, R. (2011). Racial inequality in the 21st century: The declining significance of discrimination. In *Handbook of labor economics*, volume 4, pages 855–971. Elsevier.

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, 127(3):1210–1261.

Fryer Jr, R. G. (2020). An empirical analysis of racial differences in police use of force: a response. *Journal of Political Economy*, 128(10):4003–4008.

Goncalves, F. and Mello, S. (2021). A few bad apples? racial bias in policing. *American Economic Review*, 111(5):1406–1441.

Goncalves, F. and Mello, S. (2023). Police discretion and public safety. *NBER Working Paper 31678*.

Grogger, J. and Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, 101:878–887.

Heckman, J. (1979). Sample selection bias as specification error. *Econometrica*, 71(1):53–161.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Hirschfield, P. J. (2023). Exceptionally Lethal: American Police Killings in a Comparative Perspective. *Annual Review of Criminology*, 6:471–498.

Hoekstra, M. and Sloan, C. (2022). Does race matter for police use of force? evidence from 911 calls. *American Economic Review*, 112(3):827–860.

Hull, P. (2020). Estimating hospital quality with quasi-experimental data.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jordan, A., Karger, E., and Neal, D. (2024). Early predictors of racial disparities in criminal justice involvement. Technical report, National Bureau of Economic Research.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *The American economic review*, 96(3):863–876.

Knox, D., Lowe, W., and Mummolo, J. (2020). Administrative records mask racially biased policing. *American Political Science Review*, 114(3):619–637.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3).

List, J. and Uchida, H. (2025). Socio-economic status explains one-third of the racial disparity in top academic achievement. Technical report, The Field Experiments Website.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Morris, C. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

National Academies of Sciences, E., Medicine, et al. (2022). *Reducing racial inequality in crime and justice: Science, practice, and policy.*

Neal, D. (2004). The measured black-white wage gap among women is too small. *Journal of political Economy*, 112(S1):S1–S28.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of political Economy*, 104(5):869–895.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *The Journal of Human Resources*, 23(3):279.

Newman, S. H. (2006). Proving probable cause: Allocating the burden of proof in false arrest claims under sec. 1983. *U. Chi. L. Rev.*, 73:347.

Ridgeway, G. and MacDonald, J. (2009). Doubly robust internal benchmaking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, 104:661–668.

Rivera, R. (2025). Are bad cops better police? the tradeoff between officer aggression and public safety. *Unpublished manuscript*.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002.

Rubin, D. (1981). The Bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134.

Scarborough, K. E. and Hemmens, C. (1999). Section 1983 suits against law enforcement in the circuit courts of appeal. *T. Jefferson L. Rev.*, 21:1.

Schwartz, G. L. and Jahn, J. L. (2020). Mapping fatal police violence across us metropolitan areas: Overall rates and racial/ethnic inequities, 2013-2017. *PloS one*, 15(6):e0229686.

Tapp, S. N. and Davis, E. J. (2022). Contacts Between Police and the Public, 2020. Technical report, Bureau of Justice Statistics, U.S. Department of Justice.

Weisburst, E. (2019). Police use of force as an extension of arrests: Examining disparities across civilian and officer race. *AEA Papers and Procedings*, 109:152–156.

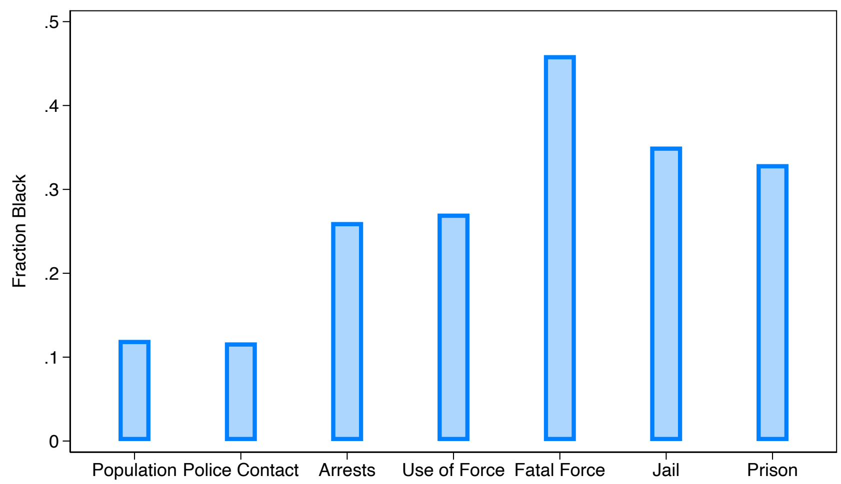
Weisburst, E. K. (2024). Whose help is on the way? *Journal of Human Resources*, 59(4):1122–1149.

West, J. (2018). Racial bias in police investigations.

Zaiour, R. and Mikdash, M. (2024). The impact of police shootings on gun violence and civilian cooperation. *Journal of Public Economics*, 237.

Zimroth, P. L. (2021). Thirteenth report of the independent monitor racial disparities in nypd stop, question, and frisk practices: An analysis of 2013 to 2019 stop reports.

Figure 1: Racial disparities at stages of criminal justice contact



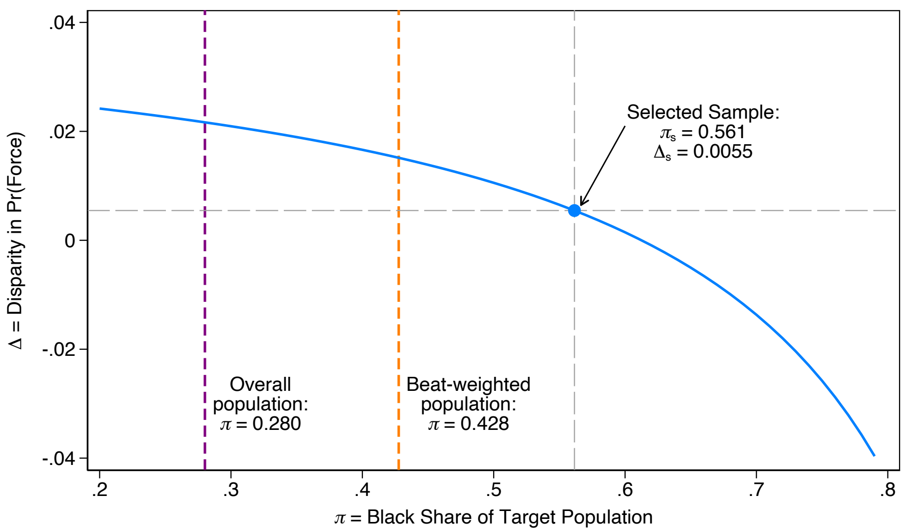
*N*otes: This figure reports the Black share of various populations facing criminal justice contact. Information on police-initiated contacts and use of force victims are from the BJS's Police-Public Contact Survey. Information on arrests are from the FBI Uniform Crime Reporting Data Program. information on Jail and Prison populations are from the BJS.

Table 1: Racial disparities in use of force among arrestees

	(1) Force	(2) Force	(3) Force	(4) Force
Race = Black	0.00600	0.00546	0.00580	0.00471
	(0.000859)	(0.00107)	(0.00108)	(0.00108)
Non-Black Mean	.023	.023	.023	.023
FE	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Crime Type	No	No	No	Yes
Arrests	134361	134361	134361	134361

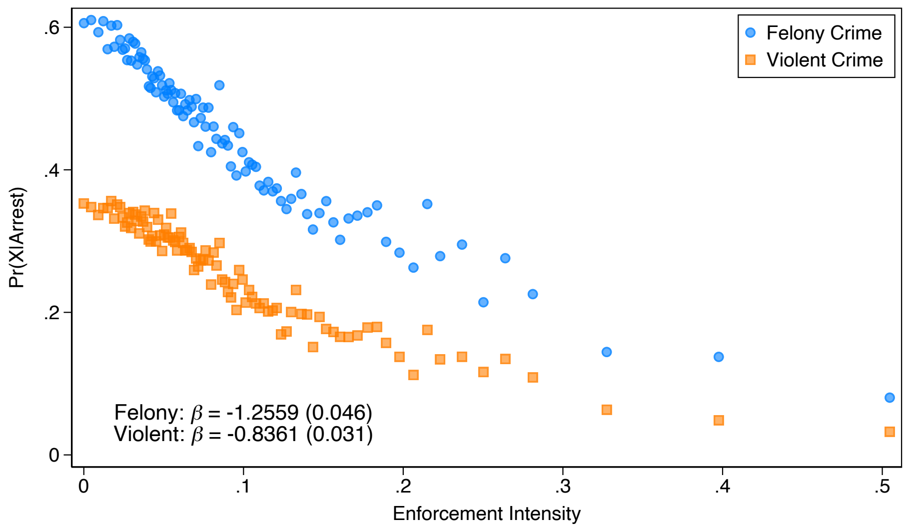
*N*otes: This table reports the racial disparities in police use of force in the selected sample of arrestees. Specifically, we regress an indicator for whether force was used during an arrest on an indicator for whether the arrestee is Black. Column 1 reports the raw disparity and column 2 adds beat × weekend × shift fixed effects as well as division × year × month fixed effects. Column 3 adds additional demographic controls for age and gender. Column 4 adds fixed effects for crime type (violent, property, drug, and other). Standard errors clustered at the assignment level in parentheses. The disparity in column 2, which conditions on assignment and time effects but not other characteristics, is the disparity we focus on adjusting for differential sample selection.

Figure 2: Sample selection adjustment curve



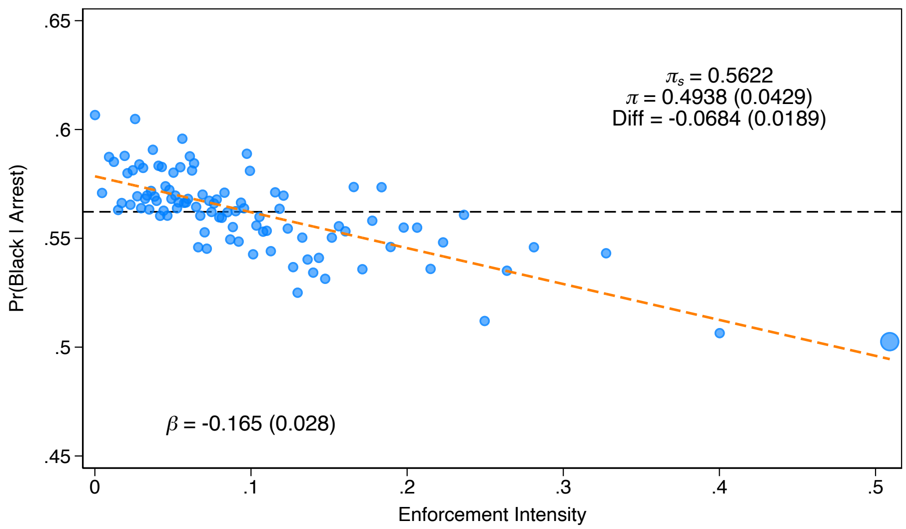
*N*otes: This figure illustrates the selection-adjusted disparity in use of force ($\equiv \tilde{\Delta}$) as a function of the racial composition of the target sample using equation 4 and substituting in the relevant moments from the selected sample. Solid blue dot denotes the moments in the selected sample. As relevant benchmarks for the composition of the target sample, we report the Black share of the overall population with the purple dashed line and the beat-weighted Black population share with the orange dashed line, computed by calculating the Black population share in each beat and then taking a weighted average, weighting by each beat's representation in the arrests data.

Figure 3: Arrestee crime type by enforcement intensity



*N*otes: This figure illustrates the relationship between the probability that an arrest is for a felony crime (blue dots) or violent crime (orange squares) crime and arrest-level enforcement intensity, computed as described in section 4.1. The figure displays conditional binscatters with 100 quantiles bins, adjusting for assignment and division-time fixed effects and reports coefficients and bootstrapped standard errors from the corresponding linear regressions.

Figure 4: Enforcement intensity and the racial composition of arrests



Notes: This figure illustrates the relationship between the probability that an arrestee is Black and arrest-level enforcement intensity, computed as described in section 4.1. The figure displays conditional binscatters with 100 quantiles bins, adjusting for assignment and division-time fixed effects and reports coefficients and bootstrapped standard errors from the corresponding linear regressions. Dashed horizontal line denotes the share Black in the selected sample, $\pi_s = 0.5622$. The vertical height of the rightmost bin corresponds to our baseline estimate of $\pi = 0.494$. Figure also reports the estimated difference between π_s and π and the associated bootstrapped standard error.

Table 2: Baseline estimates

	Racial Composition			Disparity in Force		
	(1)	(2)	(3)	(4)	(5)	(6)
	π_s	π	Difference	Δ_s	$\tilde{\Delta}$	Difference
Binscatter	0.5622	0.4938	0.0684	0.0055	0.0111	-0.0056
		(0.0429)	(0.0189)		(0.0017)	(0.0013)
Linear	0.5622	0.4952	0.0669	0.0055	0.0110	-0.0055
		(0.0429)	(0.0125)		(0.0013)	(0.0008)

*N*otes: This table reports our baseline estimates of the Black share of the target sample of potential arrestees (π) and the associated selection-adjusted force disparities $\tilde{\Delta}$, computing from the estimated π using equation 4. In both cases, we also report the difference between the unselected and unselected parameters. Bootstrapped standard errors clustered at the assignment level are reported in parentheses. The first row reports estimates from our baseline binscatter approach, while the second row reports estimates from a linear specification.

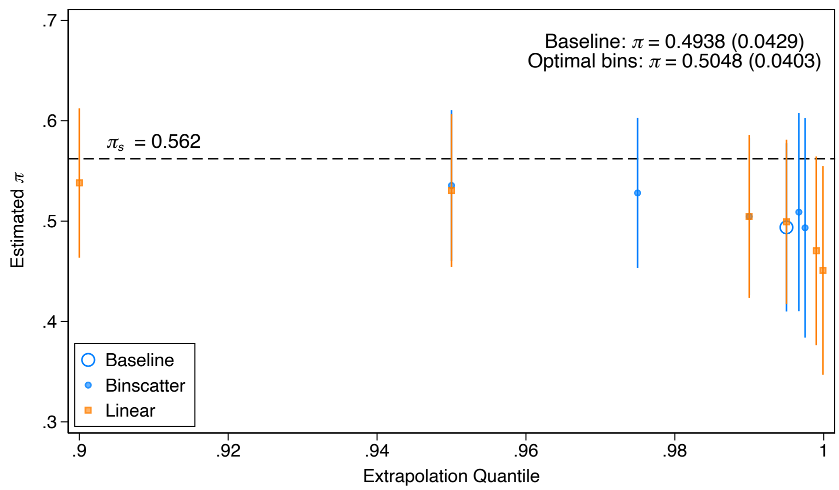
Table 3: Unscaled estimates for target sample

	Target sample estimates:		
	(1)	(2)	(3)
	Black	Non-Black	Difference
Pr(Arrest)	0.1266	0.0962	0.0304
	(0.0058)	(0.0073)	(0.0078)
Pr(Force)	0.0036	0.0022	0.0014
	(0.0003)	(0.0003)	(0.0002)

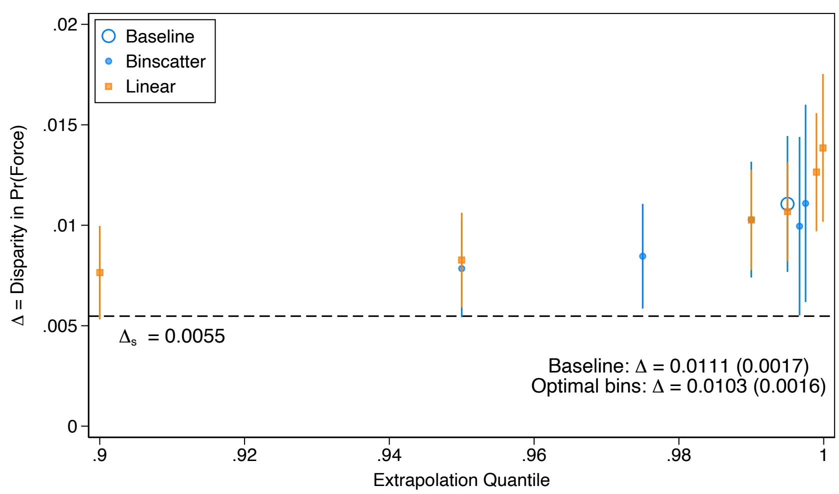
*N*otes: This table reports estimated rates of selection and force for the target sample (i.e., disparities which are not rescaled by baseline selection rates). The bottom right number (0.0014) corresponds to the Δ (rather than $\tilde{\Delta}$) described in section 2.

Figure 5: Sensitivity of estimates to extrapolation quantile

(a) Estimated π by extrapolation quantile

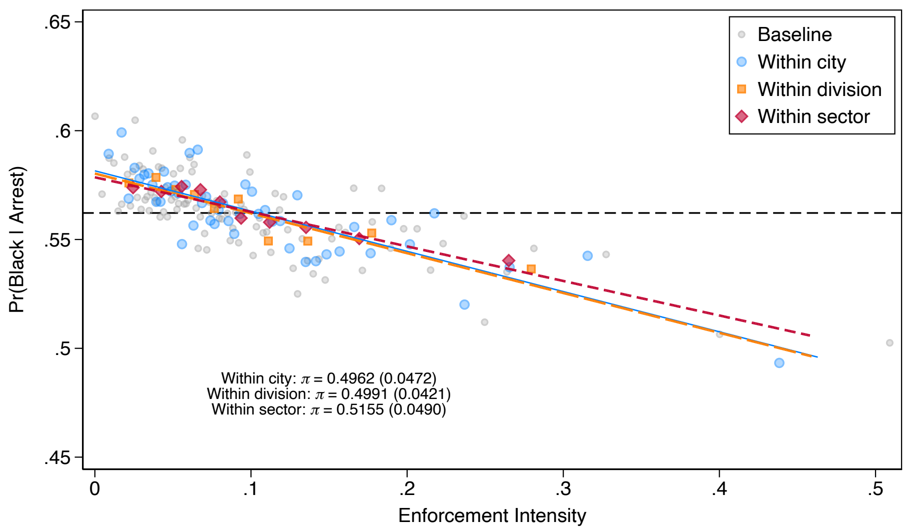


(b) Range of estimated $\tilde{\Delta}$



Notes: Panel (a) reports estimates of π and associated 95 percent confidence bands using specifications which vary the method for obtaining the estimate (binscatter and linear) and the quantile used to define the maximally enforcing officers (the horizontal axis). Panel (b) reports the selection-corrected force disparity estimates ($\tilde{\Delta}$) associated with each estimated π. Our baseline estimates, shown in figure 4, are those based on a binscatter using the top one percent of the enforcement intensity distribution as the definition of maximal. Each figure also reports the estimate when using the Cattaneo et al. (2024) data-driven optimal bandwidth selection procedure to construct the binscatter.

Figure 6: Within-location estimates of π



*N*otes: This figure illustrates the relationship between the probability that an arrestee is Black and arrest-level enforcement intensity. The smallest gray dots denote estimates from our baseline specification (same as depicted in figure 4). Blue circles, orange squares, and red diamonds report estimates from our within-location approach for different notions of location, obtained by computing location-specific enforcement intensity measures, separately estimating the relationship between enforcement intensity and the probability that an arrestee is Black for each location, and then aggregating up the location-specific estimates, weighting by location shares in the arrest data. The figure reports the associated within-location estimates of π and associated standard errors.

Table 4: Covariates by race in selected and target samples

	(1) Black	(2) Non-Black	(3) Difference
Panel A: Selected sample ($P_i = 1, S_i = 1$)			
Covariate index	0.0245	0.0232	0.0013
(Predicted force)	(0.0022)	(0.0020)	(0.0004)
Joint tests of differences:			
Demographics			$p < 0.01$
Crime type			$p < 0.01$
All covariates			$p < 0.01$
Panel B: Target sample ($P_i = 1$)			
Covariate index	0.0190	0.0189	0.0001
(Predicted force)	(0.0018)	(0.0017)	(0.0005)
Joint tests of differences:			
Demographics			$p < 0.01$
Crime type			$p = 0.38$
All covariates			$p = 0.06$

*N*otes: This table reports the race-specific distribution of a covariate index in the selected and target samples using the approach explained in section 6.1, where the index is constructed from a regression of force on non-race characteristics using only the non-Black arrestees. Panels (a) and (b) report information for the selected sample of arrestees and the target sample of potential arrestees, respectively. In each panel, we report the race-specific average covariate index (as discussed in section 6.1) as well as the Black versus non-Black difference. We also report p-values from joint tests of the null hypothesis that groups of covariates are equal across race. Covariates included in "demographics" are gender and age, parameterized as indicators for age 0-24, age 25-34, age 35-44, and age 45+, while covariates included in "crime type" are indicators for violent, property, drug, or other crime. Corresponding averages for all covariates are reported in appendix table A-12. Generically, to calculate the covariate distributions in the target sample, we first conduct our baseline estimation approach with $\mathbb{I}[X_i = x]$ on the left-hand side to identify $Pr[X_i = x]$ for all demographic cells. Then, we restrict the sample to $X_i = x$ and conduct our approach with $R_i = b$ on the left-hand side to identify $Pr[R_i = b|X_i = x]$. For the covariate index specifically, we execute this procedure for quintile bins of predicted force and compute implied values based on the bin-specific means.

Table 5: Implications of estimated selection patterns for Fryer Jr (2019) estimates

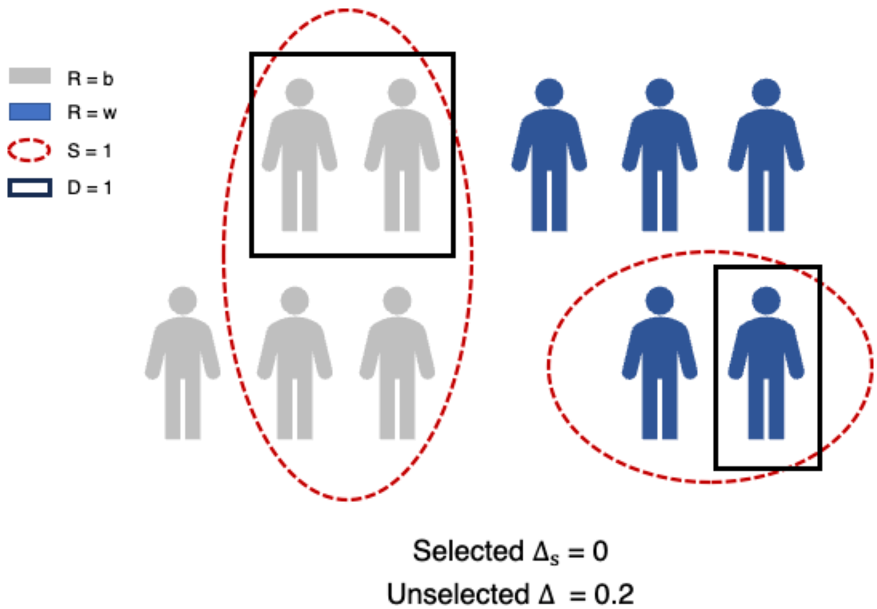
		Overall		Within-X	
	(1)	(2)	(3)	(4)	(5)
	$E(D_i \mid R_i = w, S_i = 1)$	Δ_s	$\tilde{\Delta}$	Δ_s	$\tilde{\Delta}$
NYC SQF (Any Force)	0.0080	0.0187	0.0206	0.0136	0.0141
Police-Public Contact Survey (Any Force)	0.1530	0.0817	0.1185	0.0265	0.0373
Houston: Narratives (Shooting)	0.4550	-0.1083	0.0011	-0.0346	-0.0023
Houston: Taser (Shooting)	0.1850	-0.0679	-0.0234	-0.0572	-0.0440
Houston: Arrests (Shooting)	0.1500	-0.0501	-0.0140	-0.0324	-0.0218

*N*otes: This table reports estimates from Fryer Jr (2019) and selection-adjusted estimates based on the estimated degree of differential selection in our data. Each row corresponds to a different setting and/or specification in Fryer Jr (2019). The first column reports the setting/specification-specific probability of force for white civilians in the selected sample. Columns 2 and 4 report the estimated Black-white force disparity in the selected sample with and without covariates (note that these are converted from the odds-ratios that he reports). Columns 3 and 5 report the selection adjusted estimates, computed using Equations (4) and (8). In column 3, we use our baseline estimate of π to compute the scaling ratio $\frac{\pi}{1-\pi} \cdot \frac{1-\pi_s}{\pi_s}$ in Equation (4). In column 5, we use a version of this ratio which is estimated separately within covariate cells and then averaged across cells, weighting by estimated sample shares in the unselected sample.

ONLINE APPENDIX

A Appendix figures and tables

Figure A-1: Visualizing the identification problem



*N*otes: This figure visually illustrates the central identification challenge of interest. The potentially-selected target sample is comprised of five non-Black and five Black individuals (so $\pi = Pr[R_i = b] = 0.5$). Dashed ovals denote which individuals are selected into data ($S_i = 1$) and solid rectangles denote which individuals are treated ($D_i = 1$). Four Black individuals are selected and two are treated, while two non-Black individuals are selected and one is treated. Hence, the selected sample disparity is $\Delta_s = 0$, while the unselected disparity is $\Delta = 0.2$.

Table A-1: Comparison of Chicago and Seattle to U.S. Cities

	Chicago		Seattle		> 500k Pop.		> 50k Pop.	
Total Arrests	2139.2	(372.2)	2253.1	(116.7)	3616.2	(2108.8)	3645.8	(2878.6)
Black Arrestee	4926.1	(889.3)	10492.9	(333.0)	9261.2	(6681.3)	12912.1	(12305.7)
White Arrestee	1780.4	(364.0)	1829.5	(114.5)	5233.4	(4067.8)	7524.4	(17400.9)
Index Crime Arrests	614.7	(75.4)	659.4	(27.8)	755.2	(336.2)	725.5	(381.7)
Black Arrestee	1479.3	(246.6)	3116.3	(124.0)	1994.3	(1066.0)	2937.7	(2799.4)
White Arrestee	447.4	(59.0)	533.7	(30.4)	1049.9	(765.3)	1566.8	(4575.8)
Violent	148.1	(24.6)	232.8	(19.2)	284.1	(155.7)	197.9	(132.8)
Property	466.6	(51.9)	426.7	(33.9)	471.2	(246.2)	527.6	(343.2)
Index Crimes	4158.9	(199.2)	6744.8	(173.7)	5201.6	(1713.4)	3507.6	(1626.1)
Violent	1003.6	(102.1)	689.2	(64.8)	924.6	(476.2)	454.4	(346.3)
Property	3155.4	(111.8)	6055.6	(192.6)	4277.1	(1422.8)	3053.2	(1394.7)
Use of Force - Fatal	0.66	(0.31)	0.66	(0.28)	0.59	(0.41)	0.46	(0.82)
Black Civilian	1.62	(0.53)	1.74	(2.39)	1.23	(1.62)	1.21	(6.64)
White Civilian	0.14	(0.13)	0.52	(0.39)	0.41	(0.47)	0.40	(1.38)
Number of Officers	464.3	(24.9)	217.1	(7.9)	247.8	(104.3)	159.4	(83.1)
Population	2712608		637850		1026386	(739302)	156351	(265647)
% Black	31.45		7.19		21.13	(19.95)	12.30	(15.04)
% White	32.17		66.20		40.42	(14.63)	53.34	(23.17)
% Hispanic	28.95		6.36		27.83	(18.91)	23.34	(20.92)
% Age <14	18.86		13.46		19.69	(3.12)	20.16	(3.59)
% Age 15-24	14.51		13.34		14.90	(2.04)	15.61	(5.27)
% Age 25-44	33.22		37.52		31.19	(3.49)	28.13	(3.34)
% Age >45	33.41		35.68		34.22	(2.70)	36.10	(5.71)
% Education < High School	18.37		6.88		16.90	(4.84)	14.54	(8.63)
Unemployment Rate	13.15		6.52		10.74	(3.96)	9.72	(3.41)
Poverty Rate	19.91		13.97		19.52	(4.97)	16.18	(7.65)
Median Household Income	34730		48914		35654	(8772)	40894	(14013)

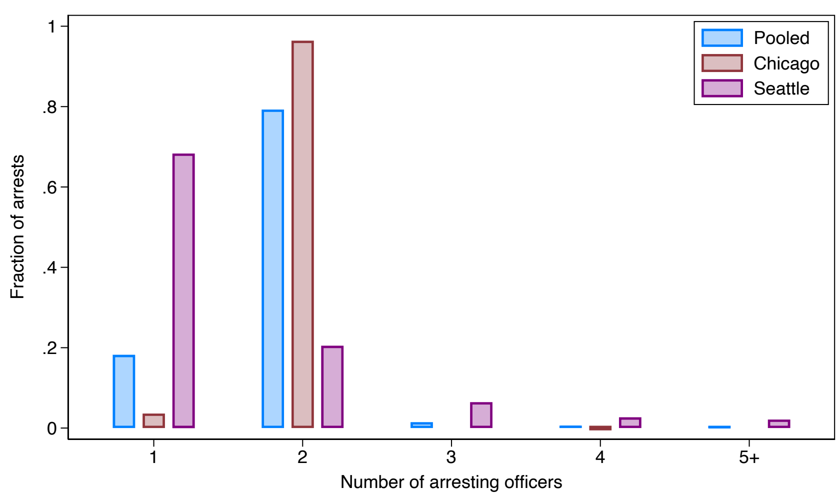
Notes: This table combines information from the FBI Uniform Crime Reports (UCR) on felony index crimes and arrests, with fatal use of force records from the crowd-sourced database *Fatal Encounters*, and demographics from the U.S. Census 5-year ACS, between 2014-2018. Arrest, crime, and use of force variables are measured as number of incidents per 100,000 residents; race specific variables are adjusted by sub-group population. Sample includes cities with complete UCR records; there are 30 (611) cities in the > 500k (> 50k) population sample.

Table A-2: Summary statistics, analysis sample of arrests

		By Race		By City	
	(1) All	(2) Black	(3) Non-black	(4) Chicago	(5) Seattle
Panel A: Arrestee demographics					
Black	0.561	–	–	0.626	0.338
Female	0.200	0.213	0.183	0.195	0.214
Age	33.63	33.46	33.85	32.89	36.18
Panel B: Arrest information					
Violent Crime	0.253	0.281	0.217	0.289	0.127
Property Crime	0.192	0.194	0.188	0.195	0.180
Drug Crime	0.074	0.079	0.067	0.085	0.034
Other Crime	0.482	0.446	0.529	0.431	0.659
Force	0.0237	0.0241	0.0233	0.0185	0.0419
N Arrests	134361	75441	58920	104041	30320

*N*otes: This table reports summary statistics for our analysis sample of arrests ($N = 134,361$).

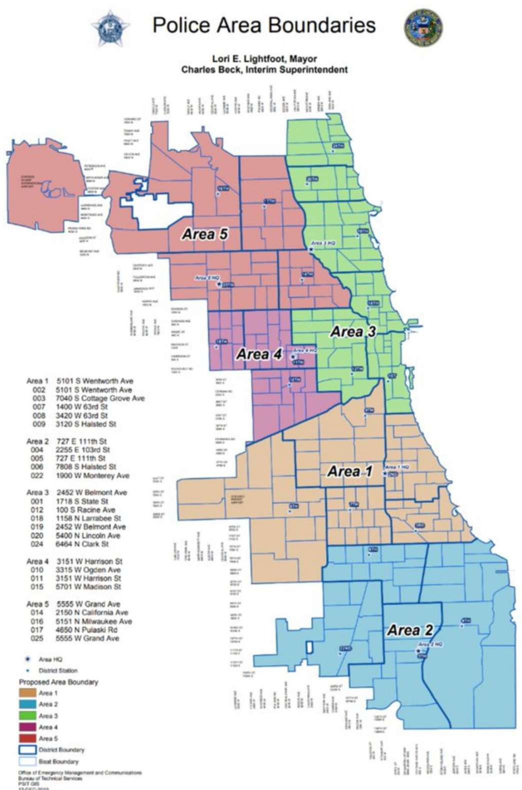
Figure A-2: Distribution of number of arresting officers



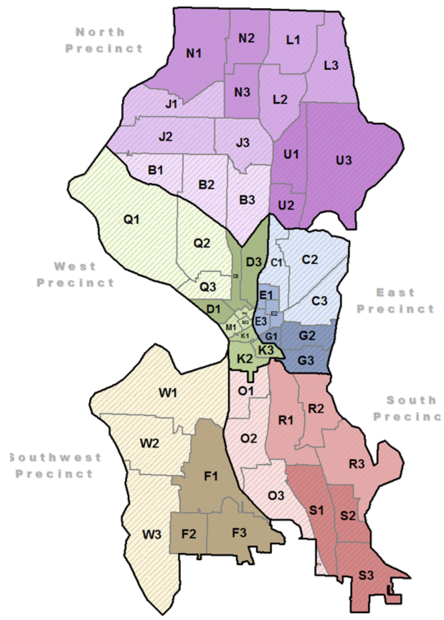
*N*otes: This figure reports the distribution of the number of arresting officers for arrests in our analysis sample ($N = 134,361$), both overall and by city.

Figure A-3: Police department geographies

(a) Chicago



(b) Seattle

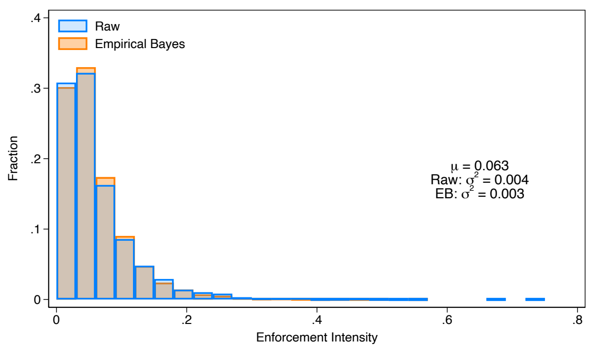


*N*otes: This figure depicts maps of the relevant police geographies for both Chicago and Seattle. As discussed in section 3.2, we adopt a harmonized terminology for geography throughout the paper where we refer to the largest areas as a "division," the second largest areas within divisions as "sectors," and the smallest areas as "beats." In Chicago, there are 5 divisions ("police areas"), 22 sectors ("districts"), and 279 beats. In Seattle, there are 5 divisions ("precincts"), 17 sectors, and 51 beats. In our baseline empirical specification, the fixed effects in our first stage estimation (equation 5) are at the level of sector × shift (time of day) × day of week and division × year × month and the fixed effects in our second stage estimation (equation 6) are the level of beat × shift × day of week and division × year × month. We explore the sensitivity of our results to changing these fixed effects in appendix table A-4.

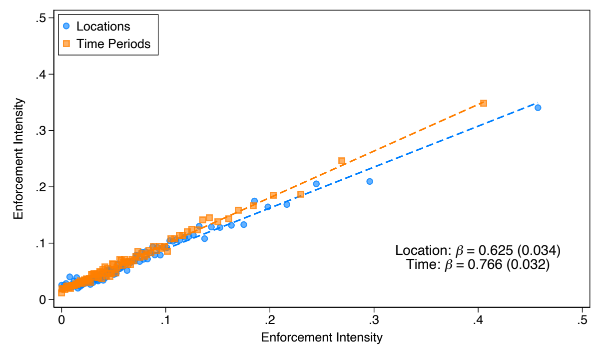
Source (Chicago): https://chicagopd.hub.arcgis.com/documents/ChicagoPD::area-district-beat-11x17-1/explore
Source (Seattle): https://www.seattle.gov/police/about-us/about-policing/precinct-and-patrol-boundaries

Figure A-4: Validating officer-level enforcement intensity

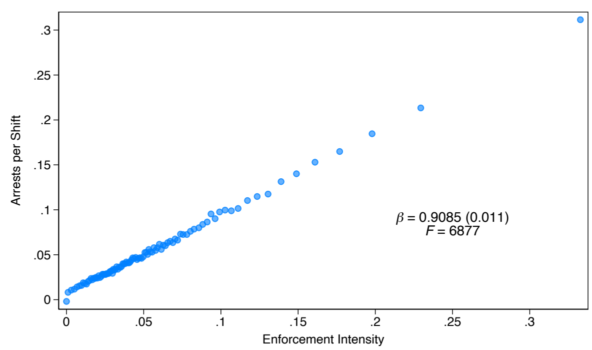
(a) Distribution of estimates



(b) Correlation of estimates



(c) First stage using partitioned instrument



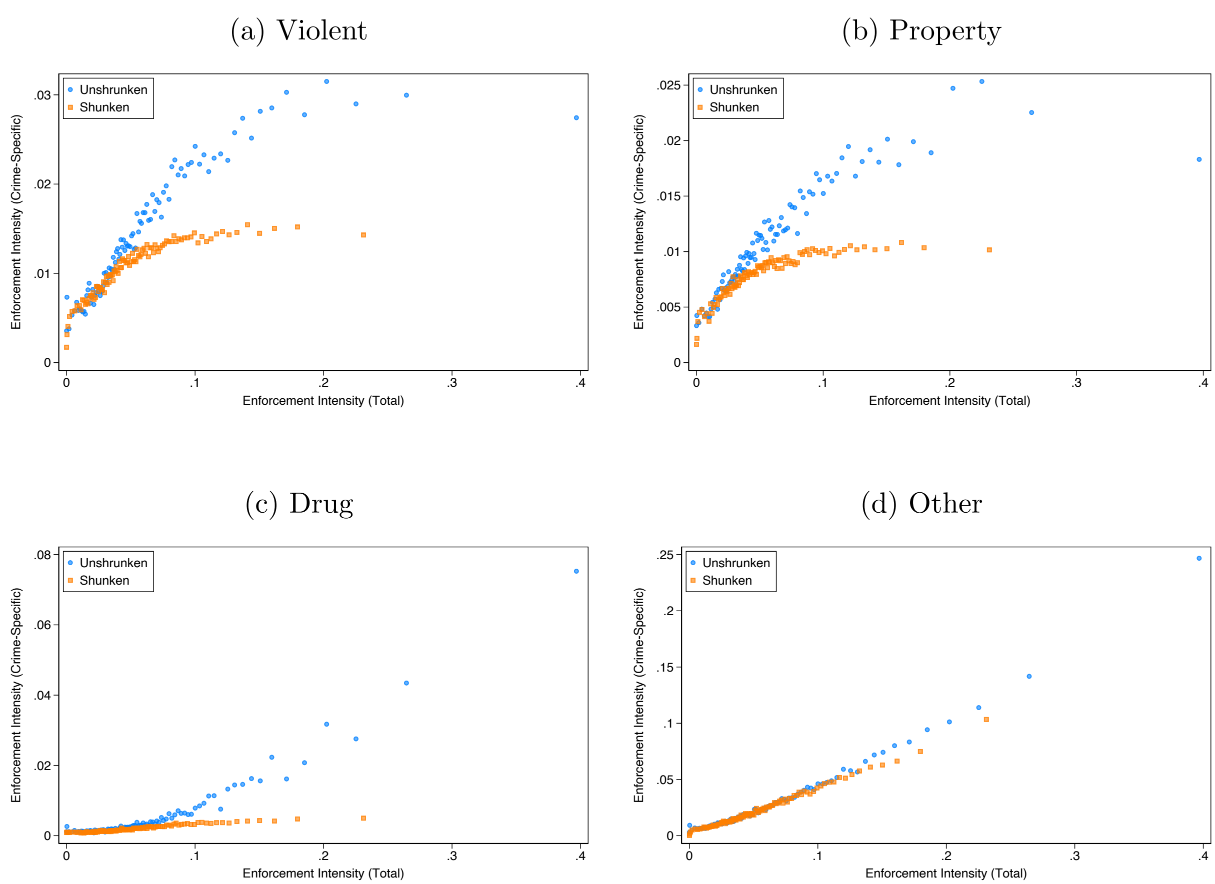
*N*otes: Panel (a) illustrates the distribution of estimated officer-level enforcement intensity (i.e., assignement-adjusted number of arrests per shift) using the method described in section 4.1. We depcit both the raw distribution and the distribution after applying empirical Bayes shrinkage (Morris, 1983) and the figure reports the variance associated with both distributions. In panel (b), we illustrate the correlation between officer $\times$ partition estimates of enforcement intensity, where the partitions are constructed based on patrol locations or time periods. In other words, this figure reports the relationship between an officer's (fixed effects-adjusted) estimated arrests per shift when working in one set of location and time periods and the same outcome when working in another set of locations or time periods. Panel (c) illustrates a "first-stage" style estimate when using our randomized partitions approach. Specifically, we randomly bin an officer's shifts into two partitions and then show the relationship between an officer's enforcement intensity in the opposite partition and their number of arrests per shift, conditional on assignment and division-time fixed effects.

Table A-3: Officer characteristics and enforcement intensity

	(1) Enforcement Intensity	(2) Enforcement Intensity
Female	-0.00366	-0.00441
	(0.00142)	(0.00136)
Race = Black	-0.00853	-0.00406
	(0.00191)	(0.00184)
Race = Hispanic	0.00127	-0.000227
	(0.00171)	(0.00160)
Race = Other	-0.00261	-0.00164
	(0.00236)	(0.00207)
Age		-0.00658
		(0.000560)
Age Squared		0.0000588
		(0.00000634)
Mean	.062	.062
City FE	Yes	Yes
Officers	4926	4926

*N*otes: This table reports regression estimates where we regress each officer's estimated enforcement intensity (the estimates presented in figure A-4) on officer characteristics, weighting by the inverse variance of the officer's estimated enforcement intensity fixed effect. Robust standard errors in parentheses. Regressions control for city fixed effects.

Figure A-5: Overall and crime-specific enforcement intensities



*N*otes: This figure depicts the relationship between an officer's overall enforcement intensity (horizontal axis) and crime-specific enforcement intensities, estimated in an identical way except using the number of arrests of a specific type as the outcome in equation (5). Blue circles depict the relationship using the raw (regression-adjusted) enforcement intensities and orange squares depict the relationship using estimates which have been shrunken via empirical Bayes. To ensure that this relationship is not mechanical, these figures use a cross-partition approach where each officer's shifts are randomly divided into two partitions, and we show the relationship between an officer's overall enforcement intensity estimated in one partition and the crime-specific enforcement intensities estimated in the opposite partition.

Table A-4: Robustness to specification choices

	Racial Composition			Disparity in Force		
	(1)	(2)	(3)	(4)	(5)	(6)
	π_s	π	Difference	Δ_s	$\tilde{\Delta}$	Difference
Panel A: First Stage Measure						
Fraction (Baseline)	0.5622	0.4938	0.0684	0.0055	0.0111	-0.0056
		(0.0436)	(0.0205)		(0.0018)	(0.0015)
Binary	0.5622	0.4924	0.0698	0.0055	0.0112	-0.0057
		(0.0414)	(0.0211)		(0.0018)	(0.0015)
Count	0.5622	0.4965	0.0656	0.0055	0.0109	-0.0054
		(0.0417)	(0.0199)		(0.0018)	(0.0014)
Count (Winsorized)	0.5622	0.4937	0.0684	0.0055	0.0111	-0.0056
		(0.0408)	(0.0197)		(0.0018)	(0.0014)
Fraction (Winsorized)	0.5622	0.4874	0.0748	0.0055	0.0115	-0.0060
		(0.0426)	(0.0199)		(0.0018)	(0.0014)
Panel B: First Stage Aggregation						
Weighted Mean (Baseline)	0.5622	0.4938	0.0684	0.0055	0.0111	-0.0056
		(0.0430)	(0.0186)		(0.0017)	(0.0013)
Unweighted Mean	0.5622	0.4913	0.0709	0.0055	0.0112	-0.0058
		(0.0430)	(0.0192)		(0.0017)	(0.0014)
Minimum	0.5622	0.5008	0.0613	0.0055	0.0106	-0.0051
		(0.0429)	(0.0204)		(0.0019)	(0.0015)
Maximum	0.5622	0.4976	0.0646	0.0055	0.0108	-0.0053
		(0.0477)	(0.0245)		(0.0019)	(0.0016)
Randomized	0.5622	0.5048	0.0574	0.0055	0.0103	-0.0048
		(0.0451)	(0.0231)		(0.0020)	(0.0016)
Panel C: Fixed Effects						
Sector in Second Stage	0.5622	0.4787	0.0835	0.0055	0.0121	-0.0066
		(0.0479)	(0.0219)		(0.0015)	(0.0014)
Combined Location and Time	0.5622	0.4734	0.0888	0.0059	0.0129	-0.0070
		(0.0479)	(0.0219)		(0.0015)	(0.0013)
Beat in First Stage*	0.5622	0.4821	0.0801	0.0055	0.0119	-0.0064
		(0.0466)	(0.0204)		(0.0014)	(0.0013)

*N*otes: This table presents estimates which are identical to the first row of table 2 except that we vary the measure used to construct enforcement intensity (panel a), the method for aggregating officer-level enforcement intensity into arrest-level enforcement intensity (panel b), and the fixed effects included in the estimation (panel c). In panel a, we winsorize the first stage measure at the 99.9th percentile. *We observe assigned beats in Chicago but not in Seattle. In this specification, we define the assignment fixed effects as City × Sector × Weekend × Shift in Seattle and City × Beat × Weekend × Shift in Chicago.

Table A-5: Split-sample IV estimates

	OLS		Split-sample IV
	(1)	(2)	(3)
	Black	Black	Black
Enforcement intensity	-0.156		-0.173
(own partition)	(0.009)		(0.007)
Enforcement intensity		-0.165	
(opposite partition)		(0.028)	

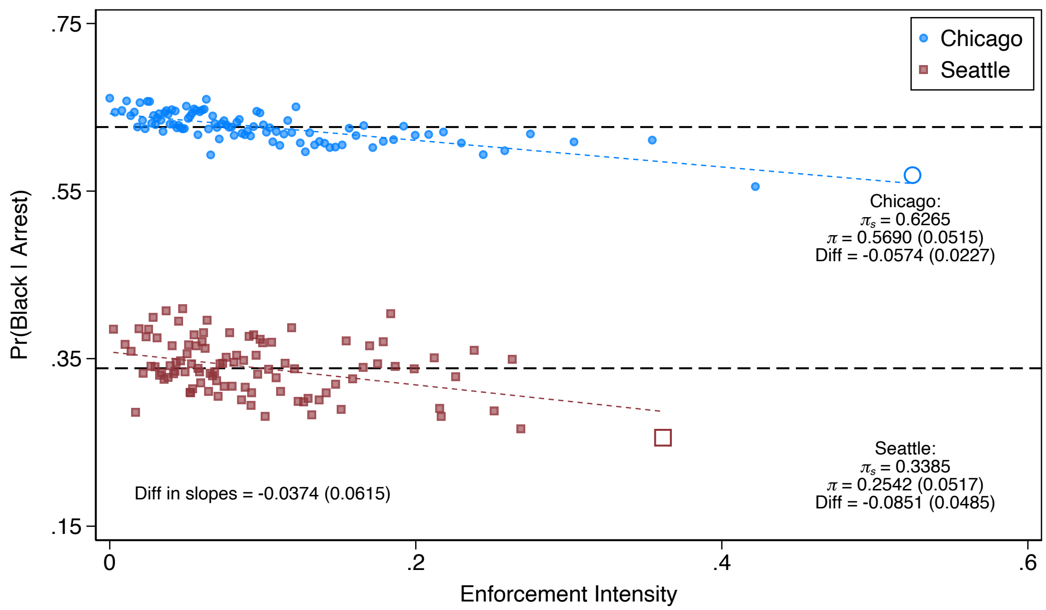
*N*otes: This table reports estimates of the linear slope from a regression of an indicator for whether an arrestee is Black on arrest-level enforcement intensity using the sample of arrestees and conditioning on assignment and division-time fixed effects. In all columns, we use the randomized partitioning of officer shifts underlying our baseline approach. In column 1, we report the estimated slope when using the arrest-level enforcement intensity in the same partition as a given arrest. In column 2, we report the estimated slope when using the arrest-level enforcement intensity from the opposite partition (this is the slope we report in figure 4). In column 3, we report a split-sample IV estimate where we instrument enforcement intensity in the arrest's partition with estimated enforcement intensity in the opposite partition.

Table A-6: Measurement error bounds on π and $\tilde{\Delta}$

	Racial Composition		Disparity in Force	
	(1)	(2)	(4)	(5)
	π_s	π	Δ_s	$\tilde{\Delta}$
Panel A: Overall estimates				
Baseline	0.5622	0.4938	0.0055	0.0111
Most conservative		0.5368		0.0077
Least conservative		0.4051		0.0164
Panel B: Within-city estimates				
Baseline	0.5622	0.4985	0.0055	0.0107
Most conservative		0.5374		0.0088
Least conservative		0.4288		0.0160

*N*otes: This table reports upper and lower bound estimates for the parameters of interest as described in section 5.2. Specifically, we estimate $\hat{\pi} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{N} \cdot \hat{q}$, where β_0 and β_1 are in the intercept and slope from a regression of whether an arrestee is Black on arrest-level enforcement intensity $\tilde{N}$, conditioning on assignment fixed effects (i.e., the relationship depicted in figure 4). Here we use the the split-sample IV estimate of the slope β_1 (reported in table A-5) and use upper and lower bound estimates of the maximum enforcement intensity $\hat{q}$ to construct most and least conservative bounds. Our most conservative bound uses the 99th percentile of the empirical Bayes shrunken distribution of the enforcement intensity measures. Our least conservative bound uses the maximum of the unshrunken distribution of enforcement intensity measures. Panel (b) reports the same information except that everything is computed separately by city and then averaged.

Figure A-6: Setting-specific estimates of π



*N*otes: Same as figure 4 except that we report the relationship separately by city.

Table A-7: Estimates dropping arrests for officer assault

	Racial Composition			Disparity in Force		
	(1)	(2)	(3)	(4)	(5)	(6)
	π_s	π	Difference	Δ_s	$\tilde{\Delta}$	Difference
Binscatter	0.5616	0.4929	0.0687	0.0050	0.0103	-0.0053
		(0.0430)	(0.0189)		(0.0017)	(0.0013)
Linear	0.5616	0.4944	0.0672	0.0050	0.0102	-0.0052
		(0.0430)	(0.0126)		(0.0012)	(0.0008)

*N*otes: Same as table 2 except that we drop arrests which are for assault on officers ($N = 565$).

Table A-8: Estimates for use of force with civilian injuries

	Racial Composition			Disparity in Force		
	(1)	(2)	(3)	(4)	(5)	(6)
	π_s	π	Difference	Δ_s	$\tilde{\Delta}$	Difference
Binscatter	0.5622	0.4938	0.0684	0.0007	0.0021	-0.0014
		(0.0437)	(0.0211)		(0.0007)	(0.0004)
Linear	0.5622	0.4952	0.0669	0.0007	0.0020	-0.0014
		(0.0428)	(0.0124)		(0.0006)	(0.0002)

*N*otes: Same as table 2 except that we focus only on force incidents associated with civilian injury. The share of force incidents in our main sample associated with civilian injury is 26 percent and the share of non-Black arrests associated with an injury-causing force incident is 0.5787 percent. In Chicago, our data indicate whether a force incident was associated with an injury. In Seattle, our data include a measure of "force level" and we use force incidents associated with the levels "force reasonably expected to cause injury exceeding transitory pain" and "force reasonably expected to cause bodily harm."

Table A-9: Moment inequality tests for extremum-officer monotonicity

	Bootstrap p-value	
	(1) Binscatter	(2) Linear
Black	0.24	0.52
Female	0.09	0.09
Age 0-24	0.18	0.29
Age 25-34	0.22	0.60
Age 35-44	0.82	0.94
Age 45+	0.43	0.36
Violent crime	0.82	0.91
Property crime	0.80	0.92
Drug crime	0.13	0.15
Other crime	0.13	0.17
Joint: all covariates	0.12	0.11

*N*otes: This table reports bootstrapped p-values for the moment inequality test for split-sample monotonicity, described in further detail in appendix B.

Table A-10: Estimates of π conditional on covariates

	(1) π_s	(2) π	(3) Difference
Baseline	0.5622	0.4938	0.0684
		(0.0429)	(0.0189)
Baseline + controls	0.5622	0.5069	0.0553
		(0.0435)	(0.0191)
Within crime type	0.5622	0.5030	0.0592
		(0.0530)	(0.0291)
Within crime type	0.5622	0.4863	0.0759
(crime-specific propensities)		(0.0484)	(0.0258)
Within covariates	0.5622	0.5090	0.0532
		(0.0503)	(0.0264)

Notes: This table reports estimates of π using specifications which also condition on other arrestee covariates. In the second row, we repeat our baseline binscatter approach except that, in the second stage, we additionally include covariate cell fixed effects at the level of gender $\times$ age bin (0-24; 25-34; 35-44; 45-54; 55+) $\times$ crime type (violent; property; drug; other). Third row reports within-crime type estimates, obtained by separately estimating π for each crime type and aggregating up the crime-type estimates by each crime type's estimated share in the target sample. Fourth row repeats the analysis in the third row except that we also use crime-specific first stage estimates (i.e., to estimate the violent crime π, we use an officer's violent-crime specific enforcement intensity). The final row conducts within-X estimation, obtained by estimating a π separately for each covariate cell (same cells as in the second row) and aggregating up, weighting by each covariate groups estimated share in the target sample. All analyses use the binscatter approach.

Table A-11: Covariates across samples

	(1) Selected sample ($P_i = 1, S_i = 1$)	(2) Target sample ($P_i = 1$)	(3) Not selected sample ($P_i = 1, S_i = 0$)
Black	0.5622	0.4938	0.4852
	(0.0348)	(0.0429)	(0.0440)
Female	0.1995	0.1780	0.1753
	(0.0035)	(0.0166)	(0.0187)
Age 0-24	0.2823	0.2715	0.2702
	(0.0069)	(0.0188)	(0.0207)
Age 25-34	0.3030	0.3373	0.3416
	(0.0042)	(0.0178)	(0.0199)
Age 35-44	0.1848	0.1964	0.1979
	(0.0044)	(0.0118)	(0.0131)
Age 45+	0.1980	0.1578	0.1528
	(0.0042)	(0.0138)	(0.0154)
Violent crime	0.2530	0.0314	0.0037
	(0.0099)	(0.0161)	(0.0178)
Property crime	0.1915	0.0432	0.0246
	(0.0118)	(0.0164)	(0.0176)
Drug crime	0.0737	0.1348	0.1425
	(0.0080)	(0.0344)	(0.0381)
Other crime	0.4819	0.7906	0.8292
	(0.0119)	(0.0494)	(0.0553)

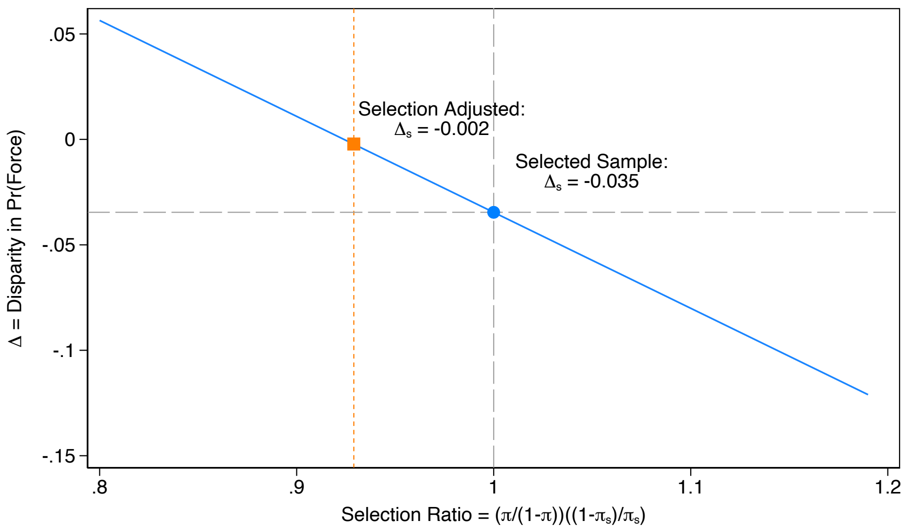
*N*otes: This table reports average covariates of arrestees/potential arrestees across samples. Column (1) reports the averages in the selected sample of arrestees. Column (2) reports the averages for the target sample, estimated using our baseline approach. Column (3) reports the averages for the not selected sample (potential arrestees who are not arrested), using the identity $Pr[X_i = x|P_i = 1] = Pr[X_i = x|P_i = 1, S_i = 1]Pr[S_i = 1|P_i = 1] + Pr[X_i = x|P_i = 1, S_i = 0]Pr[S_i = 0|P_i = 1]$. To calculate the covariate distributions in the target sample, we first conduct our baseline procedure (section 4.3) with $\mathbb{I}[X_i = x]$ on the left-hand side to identify $Pr[X_i = x]$ for all demographic cells. To calculate the probability of selection into the sample, we use $Pr[S_i = 1|P_i = 1] = E[\tilde{N}_j]/\tilde{N}_{j^*}$, as described in section 5.1.

Table A-12: Covariates by race and sample

	Selected sample ($P_i = 1, S_i = 1$)			Target sample ($P_i = 1$)			Not selected sample ($P_i = 1, S_i = 0$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Black	Non-Black	Diff	Black	Non-Black	Diff	Black	Non-Black	Diff
Covariate index	0.0245	0.0232	0.0013	0.0190	0.0189	0.0001	0.0182	0.0184	-0.0002
(Predicted force)	(0.0022)	(0.0020)	(0.0004)	(0.0018)	(0.0017)	(0.0005)	(0.0018)	(0.0017)	(0.0005)
Female	0.2127	0.1826	0.0301	0.1746	0.1814	-0.0068	0.1690	0.1812	-0.0122
	(0.0045)	(0.0052)	(0.0075)	(0.0212)	(0.0194)	(0.0194)	(0.0245)	(0.0214)	(0.0219)
Age 0-24	0.3067	0.2501	0.0566	0.3117	0.2316	0.0800	0.3124	0.2297	0.0827
	(0.0078)	(0.0097)	(0.0091)	(0.0267)	(0.0184)	(0.0228)	(0.0301)	(0.0199)	(0.0254)
Age 25-34	0.2835	0.3272	-0.0437	0.3140	0.3593	-0.0452	0.3184	0.3627	-0.0442
	(0.0041)	(0.0046)	(0.0051)	(0.0231)	(0.0171)	(0.0224)	(0.0265)	(0.0188)	(0.0254)
Age 35-44	0.1623	0.2130	-0.0507	0.1728	0.2190	-0.0462	0.1744	0.2196	-0.0453
	(0.0043)	(0.0059)	(0.0045)	(0.0152)	(0.0162)	(0.0178)	(0.0173)	(0.0179)	(0.0200)
Age 45+	0.2125	0.1789	0.0337	0.1788	0.1369	0.0419	0.1739	0.1324	0.0415
	(0.0052)	(0.0052)	(0.0061)	(0.0176)	(0.0186)	(0.0209)	(0.0201)	(0.0204)	(0.0232)
Violent crime	0.2811	0.2170	0.0641	0.0406	0.0224	0.0181	0.0057	0.0017	0.0040
	(0.0127)	(0.0113)	(0.0131)	(0.0232)	(0.0119)	(0.0126)	(0.0236)	(0.0115)	(0.0129)
Property crime	0.1941	0.1881	0.0061	0.0490	0.0375	0.0116	0.0280	0.0215	0.0066
	(0.0123)	(0.0101)	(0.0135)	(0.0182)	(0.0163)	(0.0101)	(0.0207)	(0.0172)	(0.0106)
Drug crime	0.0790	0.0669	0.0121	0.1670	0.1034	0.0637	0.1798	0.1073	0.0726
	(0.0105)	(0.0064)	(0.0077)	(0.0575)	(0.0259)	(0.0546)	(0.0649)	(0.0285)	(0.0619)
Other crime	0.4459	0.5281	-0.0822	0.7620	0.8185	-0.0565	0.8078	0.8494	-0.0415
	(0.0121)	(0.0165)	(0.0144)	(0.0467)	(0.0601)	(0.0531)	(0.0543)	(0.0656)	(0.0588)
Joint test: level differences									
Demographics			$p < 0.01$			$p < 0.01$			$p = 0.01$
Crime type			$p < 0.01$			$p = 0.38$			$p = 0.48$
Both			$p < 0.01$			$p = 0.06$			$p = 0.17$
Joint test: proportional differences									
Demographics			$p < 0.01$			$p = 0.01$			$p = 0.02$
Crime type			$p < 0.01$			$p = 0.10$			$p = 0.17$
Both			$p < 0.01$			$p = 0.05$			$p = 0.16$

Notes: This table reports covariate averages by racial group (Black and non-Black) and by sample, where "unselected" refers to the subgroup who is in the target sample but not in the selected sample. Table footer reports the p-values from joint tests that covariates are equal across race for each sample. We calculate race-specific covariate distributions in the target sample using the procedure in table 4. To calculate the characteristics among the unselected individuals in the target sample ($P_i = 1, S_i = 0$), we use the identity $Pr[X_i = x|R_i, P_i = 1] = Pr[X_i = x|R_i, P_i = 1, S_i = 1]Pr[S_i = 1|R_i, P_i = 1] + Pr[X_i = x|R_i, P_i = 1, S_i = 0]Pr[S_i = 0|R_i, P_i = 1]$.

Figure A-7: Implied adjustments for Fryer Jr (2019) estimates



*N*otes: This figure illustrates our selection correction to the estimate in Fryer Jr (2019), focusing on his analysis of officer-involved shootings in Houston in the subsample with available arrest narratives. The blue line reports the selection-corrected disparity as a function of the disparity in the selected sample and the selection ratio $\frac{\pi}{1-\pi} \cdot \frac{1-\pi_s}{\pi_s}$. The point where this ratio $= 1$ corresponds to his estimate in the selected sample. Our estimate of this ratio is depicted with a vertical orange line and the orange square shows the associated selection-corrected estimate of the disparity. These estimates are also reported in the third row of table 5.

B Technical Appendix

B.1 Moment inequality test

In Section 5.3, we present a joint test of the assumptions of our model. Under exogeneity and extremum-officer monotonicity, the arrestee composition of each officer j must satisfy the following inequalities:

$$\pi_j \tilde{N}_j \leq \pi \tilde{N}_{j^*} \leq \pi_j \tilde{N}_j + (\tilde{N}_{j^*} - \tilde{N}_j),\ \forall j \in \mathcal{J},$$

where π_j and $\tilde{N}_j$ are the regression-adjusted values of officer-specific Black composition of arrestees and rate of arrest within a shift, respectively, π is the true share black in the target sample, and $\tilde{N}_j^*$ is the maximal rate of arrest across officers.[21]

To jointly test these inequalities, we estimate $\hat{\pi}_j$, $\tilde{N}_j$,[22] $\hat{\pi}$, and we construct the following estimated moments for each officer:

$$\begin{aligned} \hat{X}_j^{lb} &= \hat{\pi}_j \tilde{N}_j - \hat{\pi} \tilde{N}_j \\ \hat{X}_j^{ub} &= \hat{\pi} \tilde{N}_{j^*} - \hat{\pi}_j \tilde{N}_j - (\tilde{N}_{j^*} - \tilde{N}_j) \end{aligned}$$

If the assumptions of our empirical design are correct, these moments should satisfy the inequalities $E[\hat{X}_j^{lb}] \leq 0$ and $E[\hat{X}_j^{ub}] \leq 0$. In finite samples, however, some inequalities may be violated even when the assumptions are correct.

We will follow the bootstrapping approach of Romano et al. (2014) and Bai et al. (2022) to conduct inference on this moment inequality test. The test statistic we will construct is a maximum of the scaled moments:

$$\hat{T} = \max\{\max_j \hat{X}_j^{lb}, \max_j \hat{X}_j^{ub}, 0\}$$

And we take the following steps:

1. Conduct Bayesian bootstrap with 100 iterations. Weights are drawn from a gamma distribution and randomized at the level of beat. Calculate $\hat{X}_{j,i}^{\cdot b}$ for each moment in each iteration, where we use the superscript $\cdot b$ to generically denote both lower and upper bound moments and i denotes the iteration.

2. Calculate

$$\hat{c}^{(1)}(1-\beta) \equiv \inf\left\{c \in \mathbb{R} : Pr\Big[\max_{j,\{l,u\}} \big(\hat{X}_j^{\cdot b} - \hat{X}_{j,i}^{\cdot b}\big) \leq c\Big] \geq 1 - \beta\right\}$$

[21] To construct π_j for each officer, we construct a dataset which is at the arrest-by-arresting officer level. We then regress an indicator for whether the arrestee is Black on the same "design" fixed effects as in our baseline approach and officer fixed effects, storing the estimated officer effects. When doing so, we weight each observation by one over the number of arresting officers, ensuring that our second stage estimation using this dataset would yield the same estimated pi.

[22] In a slight abuse of notation, we use $\tilde{N}_j$ to indicate both the true and estimated officer-level arrest frequency.

In other words, for each bootstrap, calculate $\max_{j,\{l,u\}} \left(\overline{X}_j^{\cdot b} - \overline{X}_{j,i}^{\cdot b}\right)$, and then calculate the $(1-\beta)$th percentile across iterations.

3. Calculate

$$\hat{u}_{j,i}^{\cdot b} \equiv \min \left\{ \hat{X}_{j,i}^{\cdot b} + \hat{c}^{(1)}(1-\beta), 0 \right\}$$

and

$$\hat{c}^{(2)}(1-\alpha+\beta) \equiv$$
$$\inf \left\{ c \in \mathbb{R} : Pr\Big[\max \Big\{ \max_{j,\{l,u\}} \hat{X}_{j,i}^{\cdot b} - \hat{X}_j^{\cdot b} + \hat{u}_{j,i}^{\cdot b}, 0 \Big\} \leq c \Big] \geq 1-\alpha+\beta \right\}$$

In words, for each bootstrap, we calculate $\max \left\{ \max_{j,\{l,u\}} \hat{X}_{j,i}^{\cdot b} - \hat{X}_j^{\cdot b} + \hat{u}_{j,i}^{\cdot b}, 0 \right\}$, the test statistic for that bootstrap adjusted by $\hat{u}_{j,i}$, and then calculate the $(1-\alpha+\beta)$th percentile across iterations.

4. The test "passes" if $\hat{T} > \hat{c}^{(2)}(1-\alpha+\beta)$.

The parameter $\alpha \in (0, 1/2)$ is the size of the test, and β is a tuning parameter that must satisfy $0 < \beta < \alpha$. We set $\beta = 0.01$, and to solve for the p-value of the test we loop over values α and find the largest value for which the test "passes", i.e we fail to reject the null hypothesis that all moments are satisfied.

B.2 Identification with a weaker monotonicity assumption

The moment inequalities above apply under exogeneity and extremum-agent monotonicity. In this section, we show how researchers could still place informative bounds on π in a setting with a more relaxed monotonicity assumption.

Instead of requiring that there exist an officer $j^* \in \mathcal{J}$ for whom $S_{ij*} = \max_j S_{ij}$, and thus $S_{ij*} = P_i$, we can instead assume only that the size of the target sample can not be *too much larger* than the largest individual-agent selected sample. We parameterize this requirement by a number $\epsilon \geq 0$, which limits how much the target sample can exceed the most selection-prone agent. We now define this new "$\epsilon-$monotonicity" assumption:

$$\epsilon - \text{Monotonicity:} \quad Pr[P_i = 1] \leq \max_j Pr[S_{ij} = 1] + \epsilon$$

Note that the extremum-agent monotonicity assumption is a form of $\epsilon-$monotonicity with $\epsilon = 0$. Using the definition $P_i = \max_j S_{ij}$, setting $\epsilon = 0$ requires $Pr[\max_j S_{ij} = 1] = \max_j Pr[S_{ij} = 1]$, and this latter expression is achieved by the extremum agent. For this reason, $\epsilon-$monotonicity can be interpreted as a generalization of extremum-agent monotonicity.

We continue to define $\pi = Pr[R_i = b | P_i = 1]$, though the target sample is no longer identified from the most selection-prone agent. Define $\bar{p} = \max_j Pr[S_{ij} = 1]$ as the highest observed selection probability, and let $\bar{\mathcal{J}} = \{j | Pr[S_{ij} = 1] = \bar{p}]\}$ be the set of all agents with the highest observed selection probability. Under exogeneity and $\epsilon-$monotonicity, the following inequalities hold:

$$\pi_j \bar{p} \ \leq \ \pi(\bar{p}+\epsilon) \ \leq \ \pi_j \bar{p}+\epsilon, \quad \forall j \in \bar{\mathcal{J}}$$

Our empirical strategy estimates $E[R_i = b|p_j = \max_j p_j]$ by extrapolating to this value across p_j. Since this expression is an expectation across officers with maximal observed p_j, the above inequalities hold in expectation with π_j replaced by $E[R_i = b|p_j = \max_j p_j]$. Thus, our estimation strategy can still place informative bounds on π under a relaxation of extremum-agent monotonicity.

B.3 Interpreting our estimand

In this section, we present a potential outcomes framework, and we show the conditions under which our estimands of the unconditional and conditional force gap correspond to the causal effect of race on force. We then discuss the specifics of estimating force gaps that condition on observables.

Each individual i has a pair of potential selection outcomes, $(S_i(b), S_i(w))$, which depend on whether they are Black or non-Black. In addition, they have a pair of treatment outcomes, $(D_i(b), D_i(w))$. Selection into the sample is necessary for treatment, so $S_i(r) = 0 \Rightarrow D_i(r) = 0$. The individuals' realized selection and force outcomes can be represented in terms of potential outcomes using the "switching equations", $S_i = S_i(b)\mathbb{I}[R_i = b] + S_i(w)\mathbb{I}[R_i = w]$, and $D_i = D_i(b)\mathbb{I}[R_i = b] + D_i(w)\mathbb{I}[R_i = w]$.

For simplicity, we abstract throughout from the officer encountered, which would increase the set of potential outcomes to include all possible officers. We also abstract from the set of location and division-time fixed effects used empirically, and all the discussion that follows can be thought of as conditioned on location and time.

Our first estimand of interest is the selection-corrected race gap, scaled by the selection probability for Black individuals, $\tilde{\Delta} = \frac{E[D_i|R_i=b]-E[D_i|R_i=w]}{E[S_i|R_i=b]}$. A question we address here is what assumptions are required to give this estimand a causal interpretation. To that end, we introduce the notion of unconditional unconfoundedness:

Assumption 1 (Unconditional Unconfoundedness)**.**

$$(S_i(b),\, S_i(w),\, D_i(b),\, D_i(w)) \ \perp \ R_i$$

This assumptions states that an individual's realized race R_i is independent of their potential outcomes.

Proposition 1. Under Assumption 1, $\tilde{\Delta} = \frac{E[D_i(b)-D_i(w)]}{E[S_i|R_i=b]}$.

This proposition states that, under unconditional unconfoundedness, our selection-corrected force gap represents a scaled average treatment effect of race for the target sample. This proposition can be shown very simply:

$$\tilde{\Delta} = \frac{E[D_i|R_i = b] - E[D_i|R_i = w]}{E[S_i|R_i = b]} = \frac{E[D_i(b)|R_i = b] - E[D_i(w)|R_i = w]}{E[S_i|R_i = b]} \underbrace{=}_{\text{by } A1} \frac{E[D_i(b) - D_i(w)]}{E[S_i|R_i = b]}$$

We now present an additional assumption, which is that the racial gap in sample selection only goes in one direction:

Assumption 2 (Selection Monotonicity)**.**

$$S_i(b) = 0 \;\Rightarrow\; S_i(w) = 0$$

The next proposition shows that these two assumptions guarantee that our estimand also corresponds to the causal effect of race on force *for Black arrestees*.

Proposition 2. Under Assumptions 1 and 2, $\tilde{\Delta} = E[D_i(b) - D_i(w)|R_i = b, S_i = 1]$.

We prove the statement below:

$$\begin{aligned}
\tilde{\Delta} &= \frac{E[D_i|R_i = b] - E[D_i|R_i = w]}{E[S_i|R_i = b]} \\
&= \frac{1}{E[S_i|R_i = b]}\Big[E[D_i(b)|R_i = b] - \underbrace{E[D_i(w)|R_i = w]}_{=E[D_i(w)|R_i=b] \text{ by A1}}\Big] \\
&= \frac{1}{E[S_i|R_i = b]}\Big[E[D_i(b)|R_i = b, S_i = 1]E[S_i|R_i = b] + \underbrace{E[D_i(b)|R_i = b, S_i = 0]}_{=0, \text{ by } S_i(b)=0 \Rightarrow D_i(b)=0}(1 - E[S_i|R_i = b]) \\
&\quad - E[D_i(w)|R_i = b, S_i = 1]E[S_i|R_i = b] - \underbrace{E[D_i(w)|R_i = b, S_i = 0]}_{=0 \text{ by } A2}(1 - E[S_i|R_i = b])\Big] \\
&= E[D_i(b) - D_i(w)|R_i = b, S_i = 1]
\end{aligned}$$

It is worth considering what the selection monotonicity assumption is providing in this proposition. This assumption states that, if an individual is not in the sample of arrestees when Black ($S_i(b) = 0$), they would not appear in the sample of arrestees if non-Black ($S_i(w) = 0$). This guarantees that the sample of non-Black arrestees does not contain any individuals who, if their race had been counterfactually shifted to Black, would not have been arrested. Consider if we did not assume selection monotonicity. Then, $\tilde{\Delta} = E[D_i(b) - D_i(w)|R_i = b, S_i = 1] - \frac{1-E[S_i|R_i=b]}{E[S_i|R_i=b]}E[D_i(w)|R_i = b, S_i = 0] \leq E[D_i(b) - D_i(w)|R_i = b, S_i = 1]$. In that case, the non-Black arrestees have weakly *too many* force cases relative to the right comparison for the Black arrestees, so the selection-corrected force gap understates the causal effect of race on force for Black arrestees.

B.4 Conditional force gaps

In Section 6, we decompose our force gaps into components that reflect racial force gaps among individuals with the same demographic characteristics and the force gap explained by racial differences in other demographics. We now show how, when examining gaps conditional on X, we can identify causal objects under a less strict assumption.

We now introduce the notion of conditional unconfoundedness, which states that race is independent of potential outcomes among individuals with the same value of X_i:

Assumption 3 (Conditional Unconfoundedness)**.**

$$(S_i(b), S_i(w), D_i(b), D_i(w)) \perp R_i \mid X_i$$

This assumption is a standard "selection on observables" assumption. We next show that, under this assumption, we can identify the causal effect of race on force by restricting to comparisons within observable characteristics.

For notational simplicity, we will use the following shorthands throughout this section: $D^{r,x} = E[D_i|R_i = r, X_i = x]$, $p^x = Pr[X_i = x]$, and $p^{x,r} = Pr[X_i = x|R_i = r]$.

Proposition 3. Under Assumption 3,

$$\begin{aligned} D^{b,x} - D^{w,x} &= E[D_i(b) - D_i(w)|X_i = x] \qquad \text{(B-1)} \\ \sum_x \left[D^{b,x} - D^{w,x}\right]p^x &= E[D_i(b) - D_i(w)] \\ \sum_x \left[D^{b,x} - D^{w,x}\right]p^{x,b} &= E[D_i(b) - D_i(w)|R_i = b\ ] \\ \sum_x \left[D^{b,x} - D^{w,x}\right]p^{x,w} &= E[D_i(b) - D_i(w)|R_i = w] \end{aligned}$$

Equation (B-1) is proven exactly the same as Proposition 1, and the rest of the equations apply equation 1 and the law of iterated expectations.

B.4.1 Calculation of within-X gap

We perform the following procedure to estimate the set of force gaps that are conditional on X characteristics.

First, we conduct our baseline estimation approach with $\mathbb{I}[X_i = x]$ on the left-hand side to identify $Pr[X_i = x]$ for all demographic cells. Then, for each value of $X_i = x$, we restrict the sample to this set of observations and conduct our approach with $R_i = b$ on the left-hand side to identify $Pr[R_i = b|X_i = x]$, which we denote by π_x. We denote the selected Black share for X, $Pr[R_i = b|X_i = x, S_i = 1]$, by π_s^x.

Note that, because of the reduced sample size for each demographic group, our estimates of π_x are quite noisy. To improve precision, we perform the following procedure. Letting $b_x^{(0)}$ be the initial estimated value for π_x, we calculate $\lambda_x = \frac{b_x^{(0)}}{\pi_s^x}$. We then average over all X values and construct $\lambda = \sum_x p^x \lambda_x$. Our updated value for the share Black is $b_x^{(1)} = \lambda \pi_s^x$. The intuition for this procedure is that it assumes that the "overselection" of Black individuals relative to the target sample is identical for all demographic cells, allowing us to reduce imprecision in the Black share for each cell.[23]

[23] In the absence of this procedure, if we use $b_x^{(0)}$ as our estimates of the race shares in each demographic cell, we find that 40-50 percent of the unconditional force gap is within-X rather than 70-80 percent. The reason for this lower value is because $E[\frac{\hat{\pi}_x}{1-\hat{\pi}_x} \cdot \frac{1-\pi_s^x}{\pi_s^x}] \geq \frac{E(\hat{\pi}_x)}{1-E(\hat{\pi}_x)} \cdot \frac{1-\pi_s^x}{\pi_s^x}$, where the expectation is over the estimation error in $\hat{\pi}_x$. Therefore, an increase in estimation error in $\hat{\pi}_x$ biases our estimate of Δ^x downwards.

We then estimate the following for each value of X the conditional force gap, $\Delta^x = D^{b,x} - D^{w,x}$, using the analogue to Equation 4:

$$\Delta^x \;=\; \Delta_s^x\; S^{b,x} + \big[1 - \frac{\pi_x}{1-\pi_x} \cdot \frac{1-\pi_s^x}{\pi_s^x}\big] D^{w,x}$$

B.4.2 KOB decompositions

With our estimates of within-X force gaps, we the decompose our force gaps into components that reflect racial force gaps among individuals with the same demographic characteristics and the force gap explained by racial differences in other demographics.

A well-known feature of KOB decompositions is that they are non-unique (Neumark, 1988). For each value of $\gamma \in [0,1]$, we can decompose the unconditional gap Δ as

$$\tilde{\Delta} = \sum_x \frac{\Delta^x}{E[S_i|R_i=b]}\; \big[\gamma p^{x,b} + (1-\gamma)p^{x,w}\big] + \sum_x \big[p^{x,b} - p^{x,w}\big] \frac{(1-\gamma)D^{b,x} + \gamma D^{w,x}}{E[S_i|R_i=b]} \quad \text{(B-2)}$$

where $\gamma \in [0,1]$.

We present in Section 6.1 the case where $\gamma = \pi$, where the first term is weighted by the X-composition of the overall target sample. In that case, and with Assumption 3, the first term of the decomposition captures $E[D_i(b) - D_i(b)]/E[S_i|R_i = b]$, the scaled causal effect for the whole target sample.

If we instead use $\gamma = 1$, the first term is weighted by the X-composition of the Black target sample. In that case, and with Assumption 3, the first term of the decomposition captures $E[D_i(b) - D_i(b)|R_i = b]/E[S_i|R_i = b]$. If we further impose Assumption 2 of selection monotonicity, this term also reflects $E[D_i(b) - D_i(w)|R_i = b, S_i = 1]$, the (unscaled) causal effect for the Black *arrestee* sample.

If we instead use $\gamma = 0$, and with Assumption 3, the first term in Equation B-2 reflects $E[D_i(b) - D_i(b)|R_i = w]/E[S_i|R_i = b]$, the scaled causal effect of race on force for non-Black individuals in the target sample.

Table B-1 reports the results of these KOB decompositions. The unconditional force gap is, again, 0.0111 (0.0017). For the whole target sample, the conditional force gap is 0.0073 (0.0024). For the Black target sample, the conditional force gap is 0.0078 (0.0029), and for non-Black individuals it is 0.0069 (0.0020).

In all cases, the decomposition indicates that the majority of the race gap in force is due to gaps among individuals with the same observables X_i rather than differences in the $X-$composition between Black and non-Black individuals.

In the bottom panel of the table, we probe the robustness of these findings to different choices of demographic cells. When we expand the age categories from two groups (above/below 35) to five groups, the within-X gap for the target sample increases from 0.0073 to 0.0087. When we instead coarsen the demographic breakdown to only consider cells of Felony/non-Felony-by-Young/Old, the within-X gap for the target sample increases from 0.0073 to 0.0082.

Table B-1: Decomposition of racial force gap

Unconditional Force Gap ($\tilde{\Delta}$)	0.0111 (0.0017)
Within-X Gaps	
All Target Sample	0.0073 (0.0024)
Black Target Sample	0.0078 (0.0029)
Non-Black Target Sample	0.0069 (0.0020)
Robustness for All Target Sample Within-X Gaps	
5 Age Bins	0.0087 (0.0024)
Felony x Young-Old	0.0082 (0.0022)

*N*otes: This table reports estimates of within-X force gaps, calculated for different target samples, as described in appendix B.

B.5 Comparison to selection correction approach

Our study shows how to correct for sample selection in the estimation of racial disparities in use of force. An important consideration is how our approach differs in its assumptions — and what it estimates — from a traditional selection correction approach in the spirit of Heckman (1979).

To compare the two approaches, assume that each individual-officer interaction has a *potential* force outcome $D^*_{ij} \in \{0,1\}$, but force only occurs if there is an arrest, $D_{ij} = D^*_{ij} S_{ij}$. Suppose that individual arrest satisfies a threshold crossing model,

$$S_{ij} = \mathbb{1}[\pi_{ij0} + \pi_{ij1} B_i + u_i > 0]$$

and that the potential force outcome satisfies a linear equation,

$$D^*_{ij} = \alpha_{ij0} + \alpha_{ij1} B_i + \epsilon_i,$$

where at first we allow α_{ij0} and α_{ij1} to differ by officer j.

The identification problem induced by selection into $S_{ij} = 1$ can be seen from the conditional expectation of force among arrested individuals:

$$E[D_{ij}|B_i, S_{ij} = 1] = \alpha_{ij0} + \alpha_{ij1} B_i + E[\epsilon_i | \pi_{ij0} + \pi_{ij1} B_i + u_i > 0] \tag{B-3}$$

The bias in identifying α_{ij1} arises because B_i also impacts selection into the sample. If $\pi_{ij1} \neq 0$, and u is not independent of ϵ, then the coefficient on B_i in a linear regression of force among the selected sample will be a biased estimate of α_{ij1}.

First, note that the estimands of interest in this setup are α_{ij0} and α_{ij1}, which give us the racial disparities in *potential* force rates D^*_{ij}. In contrast, the approach in our main text identifies the racial disparities in realized force rates D_{ij}. One way to think about this distinction is that our approach tells us the realized racial difference in force rates in the unselected sample, while the selection correction approach tells us the racial difference in force rates if officers were required to arrest everyone in the sample, generating an increased set of realized force rates.

Second, what assumptions are required to identify the estimand? The selection correction solution is to find a variable that impacts arrests but does not impact force. In our setting, the obvious choice for this variable is the stopping officer, whose variation in arrest propensity we use in our analysis. To utilize the officer as an *instrument* for arrests, however, requires imposing that the outcome equation is not a function of stopping officer, i.e. $\alpha_{ij0} = \alpha_{i0}$ and $\alpha_{ij1} = \alpha_{i1}$. We can then use each officer's share of arrests for Black and non-Black individuals as instruments and construct estimates $\hat{u}_i$ that have independent variation in the outcome equation (B-3). However, this crucial additional assumption, that officers do not affect realized force outcomes other than through selection S_{ij}, is implausible. We therefore view our approach as more credible.

C Legal Context

The goal of this paper is to examine race discrimination in police use of force, taking seriously concerns about selection into who is exposed to a police interaction which could lead to force. Specifically, we focus on use of force events that occur during arrest interactions, which means that we must consider the population that has the potential to be arrested when adjusting for selection into the sample. Our target sample, or unselected population, is defined as the population of individuals who face the risk of arrest, or potential arrestees.

Our notion of the target sample can also be related to the population of individuals who are *legally eligible* to be arrested. In this Appendix, we discuss the legal context for two aspects of this setting and how they may be connected to our conceptual framework and empirical approach: (1) the legal basis for an arrest interaction, and (2) the legal basis for establishing discrimination in police interactions.

C.1 Legal basis for arrest

The legal standard for an officer making an arrest of an individual is probable cause, or a reasonable basis for believing that a crime may have been committed at the time of the arrest.[24] While some arrests are made with warrants previously issued by a judge, in practice, a large fraction of arrests are made without a warrant at the discretion of patrol officers at the scene, either because an officer directly observed a crime or has probable cause to believe a crime occurred.

Importantly, while an arrest charge can be dismissed after the fact by a prosecutor or a judge, a dismissal does not typically render the initial arrest interaction as illegal or as a *false arrest*. This sequence of events is important because it means that officer decisions about who to arrest, and thus which arrest interactions have the potential to lead to force, are not likely to be responsive to later determinations about whether an arrest charge is dismissed. Case law has continued to uphold the legality of an initial arrest interaction, even when charges are later dismissed. A stark example is *District of Columbia vs. Wesby (2018)*,[25] a Supreme Court case that held that officers are protected from civil lawsuits under qualified immunity for government officials, even in cases when they lacked actual probable cause because "a reasonable officer could conclude that there was probable cause" permitting the arrests at the time. It follows that the key component of the legality of an arrest is the reasonableness standard, or the principle that an arrest may only be evaluated against whether another reasonable officer may have also made the same determination in the same circumstance (Newman, 2006).

It is likely that successful claims of false arrests are rare in practice, given that the burden of proof typically falls on the plaintiff, cases are expensive to file and pursue, and the probable cause standard is a low bar (Newman, 2006; Scarborough and Hemmens, 1999).[26] When

[24] Legal Information Institute. Cornell University. https://www.law.cornell.edu/wex/probable_cause

[25] District of Columbia v. Wesby, 583 U.S. (2018). https://supreme.justia.com/cases/federal/us/583/15-1485/

[26] Scarborough and Hemmens (1999) review a sample of US Circuit of Appeals cases and find that 22 percent of false arrest claims yielded successful outcomes for plaintiffs.

the consequences for a false arrest are low, the sole discretion of an officer applies in their decisions to arrest, meaning that officer judgment determines the selection of individuals into arrest, and into risk of potential police use of force. Moreover, claims of false arrest against officers often involve incidents of police use of force, meaning that it is vital to include all discretionary arrests when considering the sample of civilians who may be susceptible to use of force, even extreme cases of arrests that may have questionable basis of probable cause.

Because the reasonableness of any given officer is subjective, one potential way to interpret this standard is to ask whether another officer would have made a particular arrest in a given situation. In this way, our use of an "extremum officer" or maximally enforcing officer to identify the race composition of potential arrestees maps to the legal standard for arrest, if we interpret this officer as a reasonable officer "reference point" for other officers' arrest behavior.

C.2 Establishing race discrimination in use of force

The legal basis for establishing discrimination in policing interactions relates to the equal protection clause of the 14th amendment of the Constitution, which specifies that no state shall "deny to any person within its jurisdiction the equal protection of the laws." Further, the Omnibus Crime Control and Safe Streets Act (1968) requires that any agency receiving federal financial assistance (which includes nearly all municipal police departments) may not discriminate on the basis of sex or religion in addition to race, color, or national origin. Police use of force against a civilian is a form of violence used by the state, which must be justified given the circumstance, and cannot be targeted toward any particular race group.

Lawsuits alleging race discrimination in policing can take multiple forms. The U.S. Department of Justice can conduct "pattern or practice" investigations of civil rights violations by police departments, which involve assessing whether any systemic issues in a department contribute to or enable misconduct. These investigations often result in consent decree agreements which mandate specific reforms necessary for the department.[27] Alternatively, private criminal defendants can sue departments for race discrimination in policing by providing evidence of *discriminatory intent*, or actions that specifically target a race group, as well as evidence of *discriminatory effect*, or differential treatment of "similarly situated persons" of another race.[28]

Statistical evidence is often employed as one component of cases or investigations alleging race discrimination. Typically, this evidence is used to show that outcomes differ for individuals in different race groups who are considered to be otherwise similar, or are "similarly situated." In practice, this is often accomplished through regression adjustments which measure an outcome such as use of force among individuals who are stopped or arrested by police, testing for the significance of race of the suspect while controlling for various sus-

[27] Civil Rights Division, U.S. Department of Justice. (2025). "How Department of Justice Civil Rights Division Conducts Pattern-or-Practice Investigations." https://www.justice.gov/archives/file/how-pp-investigations-work/dl

[28] Civil Rights Division, U.S. Department of Justice. (2025). Section VI Proving Discrimination Intentional Discrimination. Title VI Legal Manual (updated). https://www.justice.gov/crt/fcs/T6Manual6

pect demographics and crime characteristics (e.g. Zimroth, 2021). Critically, as underscored throughout this paper, simple regression adjustments are limited by the fact that they do not account for differential selection into the sample by race.

Further, a regression-based approach to measuring race discrimination typically restricts assessments to outcomes for which all individuals in a sample are observed. For example, it is difficult to establish race discrimination in whether an individual is arrested (or stopped) by police, because there is no observable control group, or there are no administrative records of individuals who might have been arrested (or stopped) but are not. Advocates and experts often reference these types of outcome populations to benchmark population samples, comparing, for example, the race share of stopped individuals to the race share in the resident population.[29] However, these comparisons are often viewed as flawed or inconclusive, as the characteristics of individuals who are at risk of arrest (or stop) likely differ from the broader population in meaningful ways that are not directly measurable in available data. Our paper develops a practical framework to measure race disparities in use of force, which explicitly addresses this benchmarking problem, and provides a novel approach to selection adjustment.

[29] Population benchmarks are common in investigative reports of discriminatory policing; see for example: ACLU of Massachusetts. (2014). Black, Brown, and Targeted: A Report on Boston Police Department Street Encounters from 20072010. https://www.aclum.org/sites/default/files/wp-content/uploads/2015/06/reports-black-brown-and-targeted.pdf