Selection Bias and Racial Disparities in Police Use of Force*

Felipe Goncalves† Steven Mello‡ Emily Weisburst§

March 20, 2025

PRELIMINARY AND INCOMPLETE;
PLEASE DO NOT CITE OR CIRCULATE WITHOUT PERMISSION

Abstract

We examine racial disparities in police use of force. A pervasive issue in studies of policing is that the available data are selected by the police. If officers discriminate when choosing whom to select into the data, disparities measured in the observed data will suffer from selection bias. We use a conceptual framework to show how a disparity of interest in the selected sample can be corrected with an estimate of the racial composition of the unselected sample. We then develop a new approach for estimating this parameter which relies on variation in enforcement intensity across officers. Using administrative data on arrests and force incidents from Chicago and Seattle, we find that Black civilians comprise 56 percent of arrestees but only 50 percent of the unselected sample of potential arrestees, indicating a race gap in arrest rates among potential arrestees. Correcting the racial disparity in force rates in the selected sample of arrestees doubles the Black-White gap in the likelihood of facing force. Among the sample of potential arrestees, non-race characteristics are similar across racial groups, suggesting an important role for officer discrimination in driving racial force gaps.

JEL Codes: J15, K42, C10

*We are grateful to Andy Jordan, Pat Kline, Adriana Lleras-Muney, Jonathan Mummolo, Susan Parker, Andres Santos, Doug Staiger, Yotam Shem-Tov, Jonathan Tebes, Heidi Williams, and Crystal Yang, as well as seminar participants at APPAM, Dartmouth, Princeton, UC-Berkeley, University of South Carolina, and the WashU Criminal Justice conference for helpful comments.

†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

A pervasive feature of the United States criminal justice system is the presence of large disparities in outcomes across racial and ethnic groups. Black individuals are 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. The use of force by police officers is likewise uneven, with fatal and non-fatal use of force incidents more likely among black individuals.[1] These disparities translate into vastly different experiences with police across communities, with 44 percent of non-white Americans expressing confidence in police, relative to 54 percent of white Americans.[2] Motivated by these stark disparities, a large literature in the social sciences has tried to assess whether, and to what extent, these differences can be explained by racial discrimination on the part of criminal justice agents.

A central obstacle to studying racial disparities in many criminal justice settings arises from the fact that the available data include only interactions which are initiated at the discretion of criminal justice agents (e.g., Knox et al. 2020). For example, a researcher might compute the rates at which police officers use force against white and minority civilians using an administrative dataset on police stops or arrests. However, officers may also discriminate when choosing whom to stop or arrest, yielding differentially selected samples in these two groups and rendering the observed disparities challenging to interpret.

This paper measures racial disparities in police use of force when the data suffer from this type of sample selection issue. We use data from two locations, the Seattle and Chicago Police Departments, where we observe officers' work shifts, individual arrests with offense and arrestee information, and each case of police use of force. The majority of use of force cases result from interactions with an associated arrest. While the police observe or encounter many individuals for whom they have a "legal basis" for a potential arrest (a notion we formalize later), the demographics of this population is only available in cases where an arrest is made. In other words, the first recording of civilian race — the crucial variable for measuring racial disparities — occurs in data produced through discretionary enforcement.

We first present a conceptual framework that clarifies the estimand of interest and formalizes the bias induced by sample selection. Officers encounter individuals at random and can choose to arrest an individual or not; subsequently, the officer may choose to use force during the arrest. We are interested in the race gap in force outcomes among potential

[1] These gaps are visible using administrative data sources in figure 1. Our study focuses on Black vs. non-Black disparities, given both the large observed differences for this group and the salient experience of disparities among black individuals (e.g. Bass, 2001).

[2] *Gallup*, 2024. https://news.gallup.com/poll/647303/confidence-institutions-mostly-flat-police.aspx

arrestees, while we can only directly observe individuals who are arrested. We show that when Black individuals are more likely to be selected (arrested) by officers, the estimated racial disparity in use of force will be understated. We then illustrate that constructing a selection correction for the race disparity requires estimating the race composition in the unselected sample of potential arrestees. Our key insight is that this parameter of interest can be estimated using variation in enforcement behavior across officers.

Our conception of the sample of potential arrestees is based on the legal basis for an arrest. U.S. case law applies a standard whereby an arrest is valid when another "reasonable" officer would have made a similar decision in a comparable circumstance (see section 2.3). Given this legal context, we consider the set of potential arrestees as the set of individuals who would ever be arrested by any officer in our sample. For identification, we assume a monotonicity condition based on the existence of an "extreme" officer for whom, if any officer arrests an individual, the extreme officer would also arrest the individual. This condition implies that the race composition of arrestees for the extreme officer will correspond to the race composition of the sample of potential arrestees.

For estimation, we first calculate the arrest or enforcement intensity of each officer, adjusting outcomes for detailed location and time fixed effects. We then use flexible regressions to estimate the relationship between the likelihood that an arrestee is Black and the arresting officer's enforcement intensity. This relationship yields a predicted race share of arrests at an extreme quantile of officer enforcement propensity, which we then utilize as the estimate for the race share of potential arrestees.

Our primary object of interest is the difference in rates of force between Black and non-Black civilians. As in previous work, our first approach to measuring this object will be to compare force rates between Black and non-Black arrestees (as in, e.g., Fryer Jr 2019). We observe that arrests of Black individuals are 0.55 percentage points more likely to involve police force, relative to a non-Black mean of 2.3 percent, a difference that persists with controls. However, if officers differ in their arrest propensity of civilians of different races, this approach will yield a biased estimate, with that bias a function of the difference in the race composition of the selected and unselected samples.

While the Black share of arrestees is 56 percent, we estimate that the Black share of potential arrestees in the unselected sample is 49 percent. Applying our selection correction, we estimate that the true race gap in police force is 1.1 percentage points, or double the effect found in the selected sample.

These primary results are robust to alternative sample restrictions, constructions of officer enforcement intensity, treatments of cases where multiple officers are involved in an arrest,

and inclusion of granular fixed effects. Likewise, our results hold when adjusting for arrest specific controls in estimating the relationship between officer enforcement intensity and arrestee race share.

Our method relies on the monotonicity assumption that the race share of an extreme officer identifies the race share of potential arrestees. Another way to view this assumption is that each officer is choosing their set of arrestees from the set of potential arrestees that the extremum officer has arrested. This assumption implies a set of testable bounds for all other officers' race shares of arrestees. We construct a joint moment inequality test of monotonicity to test this assumption and fail to reject monotonicity in our setting.

After presenting our baseline results, we interpret and discuss our findings in the context of the existing literatures on the economics of discrimination and on racial disparities in the criminal justice system. First, using the framework of a Oaxaca-Blinder decomposition, we relate our estimand to the standard estimand in studies of discrimination (a racial disparity which is conditional on all non-race characteristics). To assess the extent to which our estimate of interest reflects discrimination as opposed to police decisions driven by other, non-race characteristics, we show that the majority of the our estimated disparity persists within (potential) arrestee characteristics, suggesting an important role for officer discrimination.

Next, we consider the implications of our findings for the conclusions in Fryer Jr (2019). This prominent study examined racial differences in police use of force across a range of data sources and force outcomes. Notably, while the study found black-white disparities for less severe forms of force, more serious types of force exhibited *negative* racial gaps. However, all the data examined arose through officer discretion (stops and arrests) and thus was subject to sample selection concerns. Assuming a proportionally similar rate of differential (by race) selection in his application and ours, we compute selection-adjusted versions of Fryer Jr (2019)'s estimates. We find that the degree of differential selection in our analysis is sufficient to erode, but not reverse, his finding of reverse discrimination in officer-involved shootings using data from Houston.

Finally, we highlight that differential arresting of potential arrestees has important implications for understanding racial disparities at later stages of the criminal justice system. We view this as an important contribution given evidence that arrests are the stage at which racial disparities in criminal justice outcomes meaningfully emerge (as we show in figure 1). As an illustration of this point, we provide a back-of-the-envelope calculation of the selection-adjusted disparity in pretrial detention using summary statistics from prior studies, finding that this selection-adjusted disparity is about three times larger than the disparity among the selected sample of arrestees.

Our paper develops a novel approach to selection adjustment which is applicable to other policing and criminal justice settings, as well as alternative settings where individual agents have broad discretion in determining inclusion into an observed sample. Such settings could potentially include caseworker decisions in foster care or homeless services, healthcare worker referrals to particular forms of treatment, or university personnel decisions about admissions. Within policing, a growing body of work has used increasingly available data on individual officers to document variation in officer behavior and discretion (e.g. Ba et al., 2021; Adger et al., 2022; Chalfin & Goncalves, 2023; Goncalves & Mello, 2021, 2023; Weisburst, 2024). Likewise, while police use of force data is not collected systematically by the U.S. government, a number of crowd-sourced databases on use of force have become available in recent years, including *FatalEncounters, Mapping Police Violence*, and *The Washington Post Fatal Force* database. These data have been used to illustrate marked differences in the share of black civilians who are victims of police use of force, relative to the black share of the population (e.g. Schwartz & Jahn, 2020). The growing prevalence of this work and data provides evidence of the applicability of our approach to other police settings.

Our work also contributes to the literature on racial disparities in police use of force within economics and political science. Fryer Jr (2019) estimates racial disparities police use of force that decrease for more serious forms of force, using a collection of settings and adjusting for detailed characteristics of interactions. This work has been the subject of critique from follow on literature that has emphasized the potential importance of potentially unrepresentative samples (Durlauf & Heckman, 2020) and selection bias in these estimates (Knox et al., 2020). Our analysis pushes beyond the critiques made in prior work by providing not only a framework for bias correction but also a direct estimate of the extent of differential selection into the sample of police encounters.

Our analysis 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 apply these ideas in pursuit of a new goal and show how these concepts can be applied in a wider variety of institutional contexts with our version of a monotonicity assumption.

Finally, our paper contributes to a new literature on discrimination in systems with multiple stages (Baron et al. 2023; Bohren et al. 2022) by illustrating the salience of this issue in a new, high-stakes setting. We show that discrimination in arrests has important implications for disparities which occur downstream from arrests and provide a framework for computing estimates which are purged of this bias.

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 $S_i = \sum_{j \in \mathcal{J}} \mathbb{1}[j_i = j] \cdot S_{ij}$ and $D_i = \sum_{j \in \mathcal{J}} \mathbb{1}[j_i = j] \cdot D_{ij}$.[3]

Individuals i differ by their race, $R_i \in \{w, b\}$, other observable characteristics X_i, and unobservable characteristics θ_i. For now, we do not examine non-race characteristics and focus on an average racial disparity in treatment.

We are interested in individuals who comprise a particular sample, whose membership is defined by $P_i \in \{0, 1\}$. For now, all we impose is that this sample weakly contains all individuals who would be selected by any officer: $P_i \geq \max_{j \in \mathcal{J}} S_{ij} \geq S_i$. 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 $P_i = 1$ as the *target sample*. For now, we keep this notion general and discuss the specific target sample of interest in detail 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 the set of all individuals in our setting.

Note that Δ does not necessarily correspond to a causal difference in treatment attributable to race. Specifically, $\Delta \neq 0$ may reflect differences in non-race characteristics which are correlated with race. We begin by focusing on this estimand to highlight how sample selection already presents issues for estimating an unconditional difference, and our first goal is to address sample selection bias in identifying Δ. We discuss the relationship between Δ and notions of discrimination in section 6.1.

[3] 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, $D_{ij} = D^*_{ij} \cdot S_{ij}$. We consider this extended formulation below when differentiating our estimand of interest from the target estimand in a traditional Heckit selection correction.

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:

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

The second equation follows from the fact that $S_i = 0$ implies $D_i = 0$. We can see that a relatively larger selection probability for minority individuals ($E[S_i|R_i = b] > E[S_i|R_i = w]$) will decrease the estimated disparity in the selected sample and can even lead to different signs for Δ and Δ_s.

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 and the latter is conditional on belonging to the larger target sample. 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, in order 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:

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

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 $\tilde{\Delta}$:

$$\tilde{\Delta} = \Delta_s + \big[1 - \frac{\pi}{1-\pi} \cdot \frac{1-\pi_s}{\pi_s}\big] E[D_i|R_i = w, S_i = 1], \qquad (4)$$

where π and π_s are the minority 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, and the challenge of identifying $\tilde{\Delta}$ reduces to identifying the minority share of the target population.[4]

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

2.3 Target sample definition

The discussion above illustrates how one can adjust the treatment disparity in the selected sample with an estimate of the racial composition of the population of interest. Of course, the implementation and interpretation of this idea depend critically on the notion of this target sample. If the target sample is the population, for example, computing $\tilde{\Delta}$ is trivial using information on population shares from the Census.[5]

Our conception of the target sample is drawn from the legal basis for an arrest, where arrest is the margin of selection into the sample, and civilians who are arrested face a risk of use of force by police officers. Specifically, we consider a target population of *potential arrestees* or individuals at risk of being arrested. In the U.S., the standard for an arrest is a reasonable basis for believing that a crime may have been committed at the time of an arrest.[6] This determination is sometimes made in consultation with a judge through a warrant but in practice is often a decision made by individual officers at the scene. Critically, while arrest charges may be dismissed by a prosecutor or judge after the fact, the legality of initial arrest decisions are rarely overturned, meaning that officer decisions about who and when to make an arrest are unlikely to respond to later dismissals of charges. In fact, the Supreme Court has held that officers are protected against civil lawsuits under qualified immunity, even in cases where they lack cause to make an arrest, as long as another "reasonable" officer would have made a similar arrest decision in a comparable circumstance.[7]

Given this legal context, our choice of target sample will be the set of individuals who would ever be arrested by any officer in our sample,

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

which we will refer to as the set of "potential arrestees." Our focus on this sample of individuals who possibly face an arrest has clear advantages over the alternative of using the overall population. Comparisons between arrested individuals and the overall population are limited by the fact that the characteristics of arrested individuals likely differ from the broader population in meaningful ways that are not directly observable; and likewise may not necessarily represent the population of interest from a researcher or public policy perspective.

[5]Note that computing Δ is also straightforward when the target sample is the population, because $Pr[S_i = 1]$ is identified by the size of the selected sample divided by the size of the population. We compute both $\tilde{\Delta}$ and Δ using population benchmarks in figure ZZ and table ZZ.

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

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

Our interest is in whether the racial disparity in use of force among arrestees is biased by differential arrests by race. Hence, we adopt the following definition of the target sample of interest: the set of individuals who are legally eligible to arrested, or in other words, the set of potential arrestees.

2.4 Approaches to Accounting for Sample Selection

Before describing our method for identifying π, we note that we are not the first to observe that differential sample selection can bias estimates of disparities in police data. We discuss here the existing set of approaches in the literature, organized through our conceptual framework.

Most close to our design and setting, Fryer Jr (2019) estimates racial differences in police use of force across multiple datasets on individuals who encounter the police, either through a police stop or arrest. He explicitly raises the possibility of sample selection bias, and his approach to accounting for this concern is to condition on a rich set of observable characteristics X measuring the civilian's demographics and the nature of the police interaction. As we describe in Section 6.2 below, this approach requires the assumption that $E[S_i|X_i, R_i] = E[S_i|X_i]$, meaning that race is not informative for selection into the sample after conditioning on X.

Another solution is to identify settings where the police do not have discretion in selecting the sample of encountered individuals. West (2018) implements this strategy in a study of racial bias in police investigations of car accidents. In his setting, officers are assigned to investigate vehicle accidents and do not have discretion to choose assigned cases. Thus, the set of recorded incidents is not a selected sample, and the observed data *is* the target sample, $S_i = P_i$.

Knox et al. (2020) present a similar conceptual framework to ours, and they also identify π as the crucial unknown parameter needed to correct for sample selection. Their suggested approach is to assume $\pi_s \geq \pi$ and place bounds on the true racial disparity based on the selected disparity.

Hoekstra & Sloan (2022) also study racial disparities in police use of force. Their approach to avoiding the sample selection issue is to target a different estimand. They observe data on all police responses to 911 calls in two cities, with information on whether the call resulted in officer use of force. While officers are quasi-randomly assigned to calls, information on the encountered individuals is only available at the point of force, which in practice generates a similar sample selection problem. To probe whether race plays a role in officer use of force, they estimate a difference-in-differences-style regression which estimates how the force rate

of white officers "scales up" when responding to calls in neighborhoods with a higher share black population, relative to the same scale up by black officers. While this approach avoids the sample selection problem, it identifies a different estimand, and it is not clear if a positive coefficient on this DiD effect implies a positive coefficient for our target estimand, or vice versa.

We also note that our estimand differs from the typical target in standard selection correction approaches in labor economics *a la* Heckman (1979). To see this, suppose that there is a potential treatment outcome D^*_{ij} in an interaction, which reflects whether force would be used *if* the individual were selected into the sample, and the realized treatment outcome is $D_{ij} = D^*_{ij} \cdot S_{ij}$. The goal of the Heckit selection correction is to estimate $\Delta^* = E[D^*_{ij}|R_i = b, P_i = 1] - E[D^*_{ij}|R_i = w, P_i = 1]$. In words, the Heckit estimand answers "what would be the racial difference in force *if all potential arrestees were arrested*." In contrast, our estimand answers "what is the racial difference in *observed force* among all potential arrestees." We discuss this distinction in more detail in Section D, and we note that estimating the Heckit estimand requires additional assumptions that are likely implausible in our setting.

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% of police interactions (over 1 million incidents) involve the threat of force or the use of force (Burghart, 2024; Tapp & 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 the 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 data on these incidents is more readily available on a national scale.

Our analysis focuses on the cities of Seattle, WA and Chicago, IL, which present a snapshot of policing in two representative large U.S. cities with varying demographics. Table A-1 shows characteristics of the sample of arrestees in both cities, and Table A-2 illustrates how these cities compare to other large and medium sized U.S. cities, using data from the F.B.I. Uniform Crime Reports (UCR) series. Chicago is one of the largest cities in the country, with 2.7 million residents, while Seattle has a population of 638 thousand. Chicago has a much higher Black resident population (31.5%), relative to Seattle (7%) and other cities

with $> 500k$ population in the U.S. (21%). Accordingly, Table A-1 shows that the share of Black arrestees is far greater in Chicago (63%), relative to Seattle (34%). Seattle also has lower rates of unemployment and poverty, and higher educational attainment and median income, than Chicago or other large U.S. cities. In terms of per capita crime, Chicago has above average violent crime rates (and accordingly more police officers per capita), while Seattle has above average property crime. Fatal use of force rates per capita are modestly higher in Seattle and Chicago than in other large cities. Across cities (including Chicago and Seattle), per capita Black arrests and fatal use of force rates against Black civilians far exceed analogous levels for white individuals. Within our analysis sample of records of individual arrests and force events in both cities, 56% of arrestees are Black, 2.4% of arrests of Black civilians result in use of force, relative to 2.3% of arrests of Black civilians (Table 1 and Figure 2).

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 among arrest events (Fryer Jr, 2019; Weisburst, 2019). A large share of non-fatal use of force incidents occur during arrests; in our sample, 62% and 80% of force events can be linked to an arrest in Chicago and Seattle, respectively.

We examine non-fatal use of force events (D_i) by police officers occurring during arrests (S_i) using data on arrestees. The sample selection problem of interest is the concern that, among the set of potential arrestees, officers may differentially arrest white and non-white civilians. Although discrimination at the arrest margin has received substantially less attention than discrimination at other stages of the criminal justice system in the literature – in part because of the absence of a clear comparison population of individuals who are not arrested but face risk of 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 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 racial disparity observed at later stages of the criminal justice system emerges at the arrest margin. Within this figure, we can also see that the race disparity in use of force appears similar in magnitude to the overall race disparity in arrests; our analysis in this paper investigates how patterns of selection into arrest may be distorting estimates of this force disparity.

3.2 Data sources

Importantly for our empirical strategy, for each setting we have "watch" data with information on each shift an officer works. These data allow us to accurately measure variation in selection propensities because we can see when officers work but do not make arrests. Watch data comes directly from the police department and includes information on officer rank, shift start and end times, and assigned geography. We link watch data to detailed individual records of arrests and incidences of use of force, using information on the officer involved and the dates and times of incidents.

Although the settings we consider use slightly different names for the various notions of police geography, we adopt a harmonized terminology for geographies throughout the paper for consistency. We use *division* to refer to the largest sub-city areas ("police area" in Chicago,"precinct" in Seattle), *sector* to refer to the largest areas within divisions ("district" in Chicago, "sector" in Seattle), and *beat* to refer to the smallest geographic division ("beat" in both settings).[8]

For our empirical work, we construct two primary datasets. The first is a panel dataset at the officer $\times$ 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 by the officer during each shift. We use only shifts worked at the rank of "police officer" and require that an officer works at least 100 such shifts for inclusion in our sample.

The second is a dataset at the arrestee level (i.e., one observation per arrest). This data set is linked to use of force records using officer information and date/time of incidents, such that it contains force outcomes. 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, and information on the officer(s) making the arrest. We can link these arrests to the first dataset at the officer $\times$ shifts using officer and date/time information.

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

4 Empirical approach

4.1 Quantifying variation in enforcement across officers

To measure variation in enforcement intensity across officers, we use our panel dataset at the officer × shift level and estimate the following regression:

$$N_{jst} = \alpha_j + \phi_s + \kappa_t + u_{jst} \tag{5}$$

where N_{jst} = the number of arrests made by officer j during a given shift,[9] the α_j's are officer fixed effects, the ϕ_s's are assignment fixed effects at the level of the assigned sector × shift (time of day) × weekend versus weekday, and the κ_t's are division × year × month fixed effects. From this regression, we compute each officer's adjusted enforcement intensity:

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

In words, $\tilde{N}_j$ is an officer's expected number of arrests per shift when working in the average assignment.

In panel (a) of appendix figure A-2, 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-2 also provides evidence that this cross-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), 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-2 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 propensity 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,

[9] 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 on 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-5.

controlling for assignment and time fixed effects.

Many arrests in our data are made by multiple officers. To construct a measure of enforcement intensity which is at the arrest, rather than officer, level, we follow the approach of Amaral et al. (2023) and take the average of the estimated $\tilde{N}_j$'s across officers credited with the arrest, weighting by each officer's total number of shifts.[10]

Figure 3 provides further validation for our approach by illustrating the relationship between this arrest-level enforcement intensity measure and the crime type associated by the 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

Having documented substantial variation across officers in their enforcement activity, we now turn to our strategy for identifying π, the minority share of the target population. Our novel approach relies on this variation in enforcement intensity across agents and builds on the "identification at infinity" ideas advanced in recent research (e.g., Hull 2020, Arnold et al. 2022).

Our approach requires two key assumptions. First, we assume that officers are randomly matched to encountered individuals, so that the characteristics of individuals, as well as their potential outcomes, are independent of their encountered officer. Second, we assume that there is an "extreme" officer for whom, if any officer arrests an individual, the extreme officer would also have arrested this individual:

1. *Exogeneity*: $(R_i, X_i, \theta_i, \{S_{ij}, D_{ij}\}_{j\in\mathcal{J}}) \perp j_i$.
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 part of the pool of potential arrestees, since $S_{ij^*} \leq \max_{j\in\mathcal{J}} S_{ij}$. Further, by our monotonicity assumption, all individuals selected by any officer would be selected by j^*, so $\max_{j\in\mathcal{J}} S_{ij} \leq S_{ij^*}$. Therefore, the extremum officer's pool of (hypothetically) selected individuals *is* the target sample of all potential offenders, $S_{ij^*} = \max_{j\in\mathcal{J}} S_{ij} = P_i$. Further, because of random assignment of officer-civilian encounters, individuals who are observed as arrested by officer j^* are in expectation identical to

[10] 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 in table A-5.

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.

Note that, while the usage of officer enforcement activity in our setting is similar in spirit to the examiner instrumental variables design (e.g. Kling, 2006; Chyn et al., 2024), our approach requires fewer assumptions. In particular, there is no exclusion restriction requiring that the encountered officer affects force only through their decision of whether to make an arrest.[11] As we discuss in Section D, 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 of π

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, let $Pr[R_i = b|S_i, N_j]$ be the share black of arrestees among officers with average enforcement activity N_j, Our estimate of π will be the value of this relationship at $N_{j^*} = \max_{j \in \mathcal{J}} N_j$:

$$\pi = \lim_{N_j \to N_{j^*}} Pr[R_i = b \mid S_i, N_j]$$

Intuitively, we want to 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 identity of the arresting officer(s). We then want to rely on that variation to estimate $Pr[R_i = b|S_i = 1, j_i = j]$ when $N_j = N_{j^*}$. 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.

In terms of implementing this idea in practice, the central complication we face is that our enforcement intensity measure captures variation in strata-adjusted arrest volume. Under

[11] In the language of Section 2.4, an exclusion restriction assumption would require $D^*_{ij} = D^*_i$ and thus $D_{ij} = D^*_i \cdot S_{ij}$.

the assumption that, within-strata, officers face the same pool of potential arrestees, a higher rate of arrests implies a higher selection propensity, but additional assumptions are necessary to rescale our enforcement intensities into selection probabilities.

Our baseline approach is to assume that extreme quantiles in the distribution of $\tilde{N}_j$ correspond to a selection probability of one. In other words, we assume that maximally enforcing officers arrest all potential arrestees. Echoing the discussion in section 2, this assumption corresponds well with the notion of the target sample as the set of potential arrestees given the legal standard that the hypothetical behavior of other officers is used to determine whether an arrest is valid. In section 5.3, we interpret our main findings under alternative assumptions about the mapping between N and p.

Given this assumption, our empirical goal becomes the identification of the Black share of arrestees among the maximally enforcing officers. We operationalize this with regressions of the following form, estimated at the arrestee-level:

$$\mathbf{1}[R_{ist} = b] = f(\tilde{N}_{ist}) + \phi_s + \kappa_t + \epsilon_{ist} \tag{6}$$

Here $\tilde{N}_{ist}$ is the arrest-level enforcement intensity constructed by aggregating the officer-level measures (as described in section 4.1), and ϕ_s and κ_t are assignment and time fixed effects as above.[12] From this regression, we can estimate π as the expected probability that an arrestee is Black at extreme quantiles of $\tilde{N}_i$.

In practice, our preferred approach is to estimate the regression in (6) non-parametrically using a conditional binscatter (Cattaneo et al., 2024), taking the fixed-effects adjusted average Black share in the top one percent of the $\tilde{N}_i$ distribution as our estimate of π. We also consider linear specifications of equation (6) and assess the senstivity of our results to using alternative quantiles below.

To avoid issues arising from the reflection problem and/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 $\tilde{N}_i$ used in the second stage regression.

[12] In our baseline approach, we replace sectors with the narrower notion of beats in the assignment fixed effects in the second stage regressions. This choice provides modest precision gains but has no impact on point estimates, as shown in appendix table A-5.

4.4 Estimating parameters of interest and statistical inference

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.

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 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.[13]

As discussed in section 2, we can also 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 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

[13]Table 2 also reports results using a linear specification for the second stage regression, extrapolating to the midpoint of the maximmal bin. Unsurprisingly given the patterns in figure 4, these results are nearly identical to those obtained using the bincsatter approach.

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 A-4 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-5, we explore the sensitivity of our baseline (binscatter) estimates to our specification choices. In panel (a), we report results when using alternative measures of N 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 the sector instead of the beat in the second stage (i.e., so that the fixed effects in the two stages are identical) and combining our assignment and time fixed effects together. This table illustrates clearly that none of these choices have any bearing on our empirical conclusions.

In appendix table A-6, we report estimates which adjust the officer-level enforcement intensity measures for estimation error using empirical Bayes methods. In other words, after estimating the officer-level enforcement intensity measures, we shrink them and then use those shrunken estimates to construct the arrest-level instrument and estimate π. While this approach slightly reduces the gap between the selected π_s and the estimated unselected π, overall conclusions are largely unaltered.

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 ($\pi = 0.4929$).

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 ($\pi = 0.538$), 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 & Miller (2022) and Goncalves & 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-3 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.

Separately, an additional concern with our setting is that some arrests involving use of force initiate with civilian choices to assault officers, rather than officer choices to arrest civilians. This potential issue is most likely to arise in arrests of civilians for officer assault, and we test the importance of this concern by removing these arrests in table A-7 from the

analysis. This exercise shows that the results are nearly identical after removing these arrest types.

5.3 Testing assumptions

As introduced in section 4.2, our approach of identifying the race composition of potential arrestees through 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 race composition of arrestees for each officer in the sample. The logic of the bounds are that the arrest rate of each officer impose limits on how much their sample composition can differ from the extremum officer. For example, an officer who arrests 99% as often as the extremum officer must have a share black among arrestees that lies between $\pi \times 0.99$ and $\pi \times 0.99 + 0.01$, the lower (upper) bound assumes that the "missing" arrestees are all black (white) civilians. Formally, defining π_j to be the fixed-effects-adjusted estimate of officer j's share black among arrestees, and p_j to be their probability of an arrest, this monotonicity assumption implies

$$\pi_j p_j \leq \pi p_{j^*} \leq \pi_j p_j + (p_{j^*} - p_j), \ \forall j$$

We can use these bounds to employ a joint test of these implied inequality conditions for all officers in the sample to test the "extremum-officer" monotonicity assumption. We do so using the inference procedure developed in Romano et al. (2014) and Bai et al. (2022). Table A-8 displays the results of this test separately for different X characteristics in the data. The first row reports the test using the share of Black arrestees (primary characteristic for this study, represented by π), and shows that the collective bounds for officer arrestee race shares satisfy monotonicity. The remaining rows repeat this test using different arrestee characteristics and nearly all of these tests likewise fail to reject monotonicity.

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. 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 Oaxaca-Blinder decomposition:

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

where $D^{r,\theta} = E[D_i|R_i = r, \theta_i = \theta]$ and $\theta^{r,\theta} = Pr[\theta_i = \theta|R_i = r]$. 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 offending.

The above decomposition also highlights that, if the race-specific distributions of θ are equal in the target sample (i.e. if, $\forall\theta$, $\theta^m = \theta^w$), then our estimated Δ corresponds to 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 characteristic means through Baye's Rule: $Pr[X = x|R = r] = Pr[R_i = r|X_i = x]Pr[X_i = x]/Pr[R_i = r]$. Our extrapolation 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 extrapolation exercise within discrete values of X. For additional details on this approach, see appendix C.

Using this approach, table 3 reports the race-specific average X's, as well as the cross-race difference, in both the selected and target sample. In the first row, we report estimates for a covariate index capturing predicted use of force.[14] In the selected sample of arrestees, there is a statistically significant gap in this covariate index: Black arrestees are predicted to face force about 0.0015 percentage points (six percent) more often based on their non-race covariates. In the target sample of potential arrestees, this gap is an order of magnitude smaller and no longer statistically distinguishable from zero.

The additional rows in table 3 tell a similar story when examining individual covariates. Specifically, gender and crime type differences by race are significantly less pronounced in

[14]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).

the target sample of potential arrestees than in the selected sample (there is no cross-race difference in age in either the selected or target sample). These estimates provide initial evidence that our unconditional force disparity is unlikely to be due to racial differences in non-race characteristics in the target sample.

To provide a more direct assessment of the relative importance of discrimination versus differences in non-race characteristics, we perform our selection-correction exercise within sets of observable characteristics. 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 the same extrapolation exercise to identify $Pr[R_i = b|X_i = x]$ and use it to estimate $\tilde{\Delta}_x$. 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 0.008 (0.002). This figure is 73% of our unconditional force disparity, 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.

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 examined racial differences in police use of force across a range of data sources and force outcomes and presented regression analyses accounting for a rich set of controls for civilian and encounter characteristics. While Fryer Jr's analyses found large racial disparities for less severe forms of force, there was no statistical difference in levels of severe or fatal force, and point estimates for these outcomes actually indicated greater force rates against *whites* in the sample.

As with the rest of the policing literature, a key limitation of the available data sources was that they all originated from incidents that involved police decision-making, such as pedestrian stops and 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 & 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 race 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, and we denote it by $\Big[\overline{\frac{\pi^{\cdot}}{1-\pi^{\cdot}} \cdot \frac{1-\pi_s^{\cdot}}{\pi_s^{\cdot}}} - 1\Big]$.[15] With this simplification, and rearranging terms, we arrive at an estimable expression for the average force disparity,

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

The first term on the right-hand side represents Fryer's estimates of race on force in the selected sample. In the second term, we take $\sum_x \tilde{\omega}_x D_s^{w,x}$ from Fryer's summary statistic measure of force against whites in the selected sample, and we construct the ratio of unselected and selected race shares from our analysis.

[15] When we estimate π_s^x separately for felony and non-felony offenses, we fail to reject that these ratios are different between the two categories, lending credence to our assumption of a single value for this ratio across covariates.

We report the results from this exercise in table 4. 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 their (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 the average force rate is higher in the second row (column 1), these coefficients increase by more. We find that the selection-corrected force gaps in the PPCS are about 40% greater than suggested by the selected sample.

The bottom three rows present coefficients from Fryer Jr's analysis of police shootings in Houston. Each row presents a separate choice for the sample of analysis. All coefficients are negative, indicating that black individuals in the sample are less likely to be shot by the police. While our corrections increase all coefficients, in only one of the six coefficients is the value now positive and still quantitatively small. To describe these findings visually, figure A-6 repeats our figure 2 but depicts the selected sample moments from row 3 of table 4. Correcting for sample selection shifts the estimates leftwards along the selection adjustment curve. The y-axis shows the original and selection-corrected values for the racial force gap, again showing lessened reverse discrimination but no evidence of a positive black-white force gap.

The conclusion from this exercise is that, while our selection correction is sufficiently important to reduce the presence of reverse discrimination observed for police shootings, it does not *reverse* the conclusions and result in large positive black-white force gaps.

6.3 Implications for other criminal justice margins

While our analysis focuses on computing the racial disparity in police of force which is purged of differential selection, we also note that our conclusions about this differential selection into arrests have implications for understanding 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, which 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, respectively, support the interpretation that

differential arrest rates by race reflect discrimination by officers, at least to some extent.

As an example to explore the implications of this discrimination at the arrest stage for understanding disparities in downstream criminal justice outcomes, we repeat our selection correction procedure using estimates from the literature which capture racial disparities at the stages immediately following an arrest. Note that this exercise re-highlights two important features of our analysis: (i) our framework applies more generally to outcomes which are conditional on selection; (ii) our selection-corrected disparity can be calculated easily using information generally reported in summary statistics tables or popular press articles for varying degrees of (assumed) differential selection.

As an illustrative example, we borrow summary statistics from Dobbie et al. (2018)'s analysis of bail hearings. In 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 these other settings would imply a target sample composition of potential arrestees $\pi = 0.512$ 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.

7 Conclusion

Understanding race discrimination in the criminal justice system is complicated by the fact that criminal justice agents have discretion in the application of sanctions at multiple stages of this system. This discretion creates the potential for differential selection into this system by race, which can distort measures of realized race disparities.

We study race discrimination in police use of force against civilian arrestees. This setting is of critical policy and public interest, as faith in the legal system can be materially impacted by community perceptions of excessive use of force by state actors. However, correctly measuring the race disparity in force requires understanding who faces the risk of use of force, and whether the choices that officers make about which civilians they detain - and the race of these civilians - matters for this disparity.

We develop a selection correction approach to estimate the disparity in use of force among potential arrestees. This approach exploits officer level arrest data to estimate the race composition of an extremum officer, under the assumption that the sample of arrestees for this officer is an approximation of the set of potential arrestees. This assumption requires that this officer's observed arrests encompass all possible potential arrests of officers who arrest fewer individuals, an assertion that we validate using a test of split-sample monotonicity. The design is grounded the legal basis for an arrest, which asserts that an officer's arrests are valid

when another reasonable officer would make a similar arrest in a comparable circumstance.

Our findings illustrate that selection matters in this setting. Using data from Seattle and Chicago, we first show that the selected sample of arrested individuals is 56% Black, and police are 24% more likely to use force on Black arrestees. However, the sample of potential arrestees is 49% Black, implying that Black civilians are more likely to be arrested in the first place. Applying our sample selection adjustment, the race disparity in use of force doubles in size - indicating that Black civilians are 48% more likely to face force.

This stark result highlights the potential for selection to affect the interpretation of race discrimination in multiple settings within the criminal justice system. Given the lack of availability of an observable control group, much of the prior literature has ignored the potential for differential selection into 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 of sanctioned individuals at later stages of the justice system.

References

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

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

Anwar, S. & 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., & Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, *112*(9), 2992–3038.

Ba, B., Knox, D., Mummolo, J., & 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., & 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., Emanuel, N., Hull, P., & Ryan, J. (2023). Racial discrimination in child protection. *NBER Working Paper*.

Bass, S. (2001). Policing space, policing race: Social control imperatives and police discretionary decisions. *Social Justice*, *28*(1 (83)), 156–176.

Bohren, J., Hull, P., & Imas, A. (2022). Systemic discrimination: Theory and evidence. *NBER Working Paper 28920*.

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

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

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

Dobbie, W., Goldin, J., & 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. & 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.

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

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

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

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. & 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.

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

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

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

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.

Romano, J. P., Shaikh, A. M., & 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. & Hemmens, C. (1999). Section 1983 suits against law enforcement in the circuit courts of appeal. *T. Jefferson L. Rev.*, *21*, 1.

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

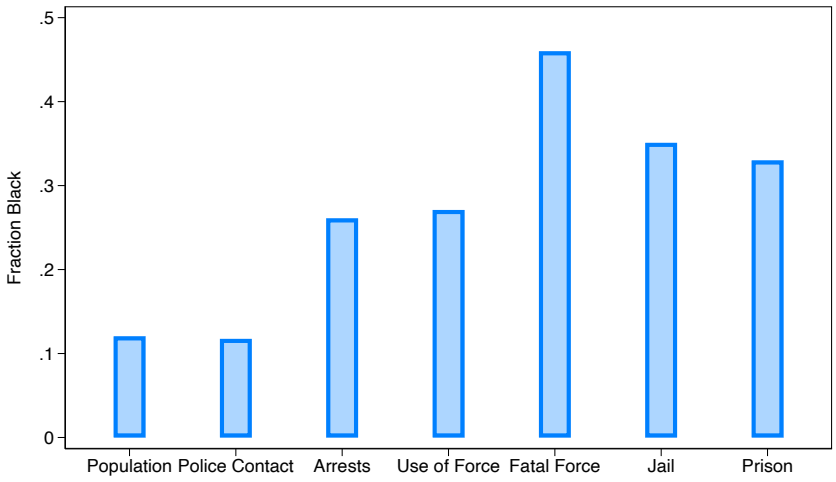
Weisburst, E. (2019). Police use of force as an extension of arrests: Examining disparities across civilian and officer race. *AEA Papers and Procedings*, *109*, 152–156.

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.

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



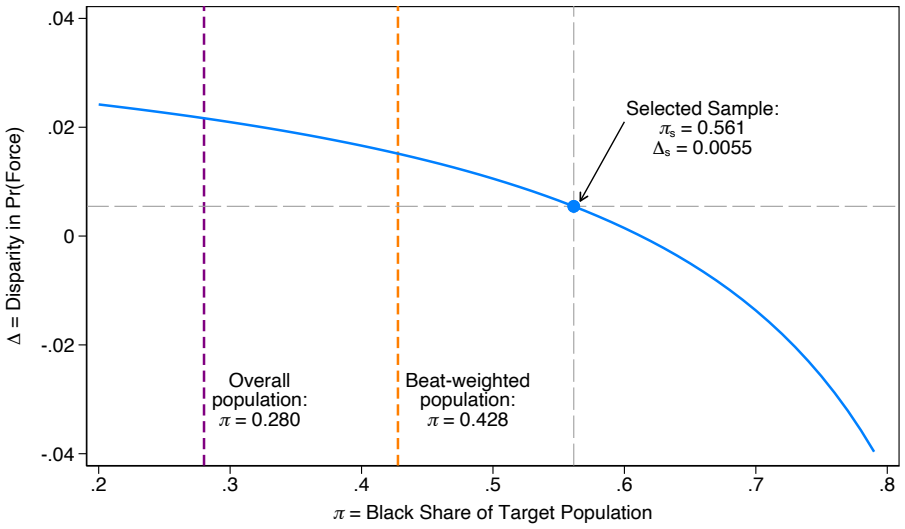
*N*otes: This figure reports the Black share of various populations facing criminal justice contact. Information on police-initiated contacts and use of force victims are from the BJS's Police-Public Contact Survey. Information on arrests are from the FBI Uniform Crime Reporting Data Program. information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

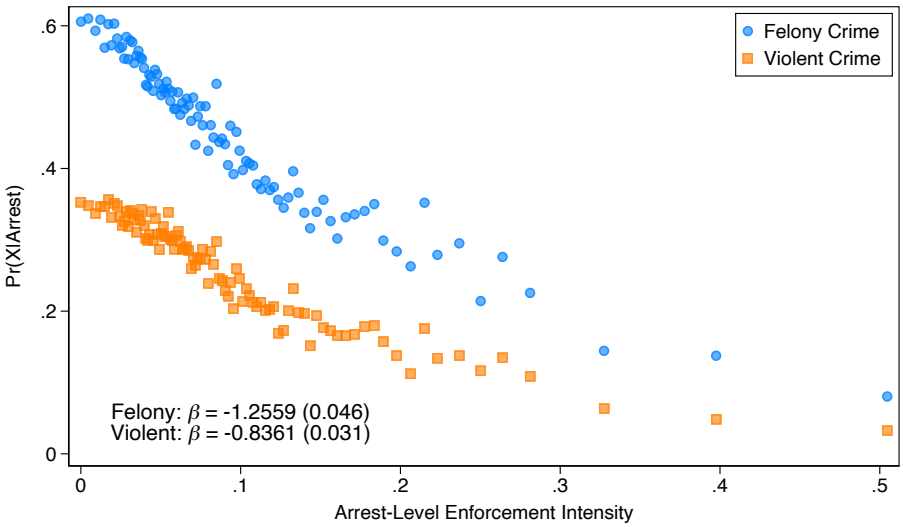
*N*otes: This table reports the racial disparities in police use of force in the selected sample of arrestees. Specifically, we regress an indicator for whether force was used during an arrest on an indicator for whether the arrestee is Black. Column 1 reports the raw disparity and column 2 adds beat $\times$ weekend $\times$ shift fixed effects as well as division $\times$ year $\times$ 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



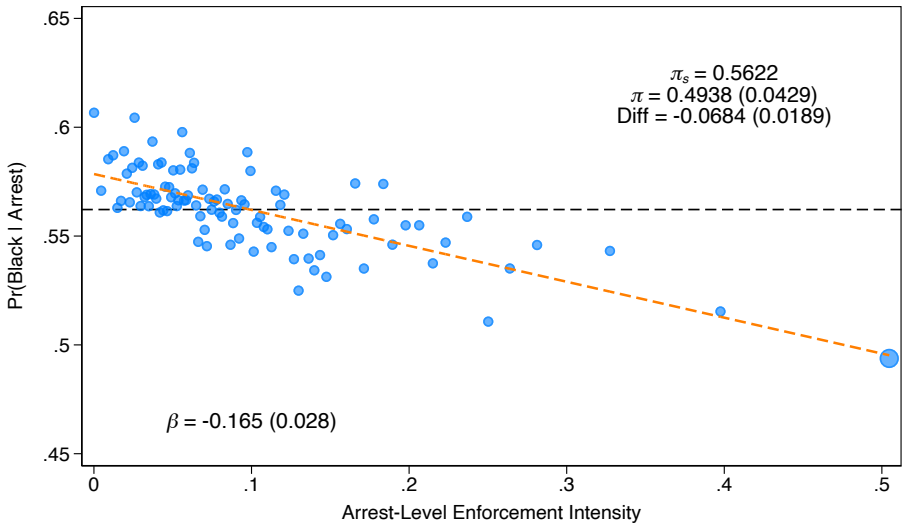
Notes: This figure illustrates the selection-adjusted disparity in use of force ($\equiv \tilde{\Delta}$) as a function of the racial composition of the target sample using equation 4 and substituting in the relevant moments from the selected sample. Solid blue dot denotes the moments in the selected 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 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 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.

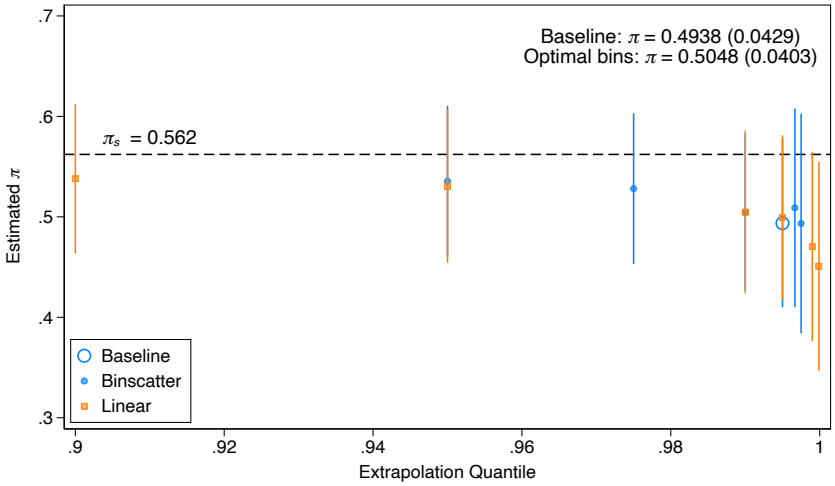
Table 2: Baseline estimates

	Racial Composition			Disparity in Force		
	(1) π_s	(2) π	(3) Difference	(4) Δ_s	(5) $\tilde{\Delta}$	(6) 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)

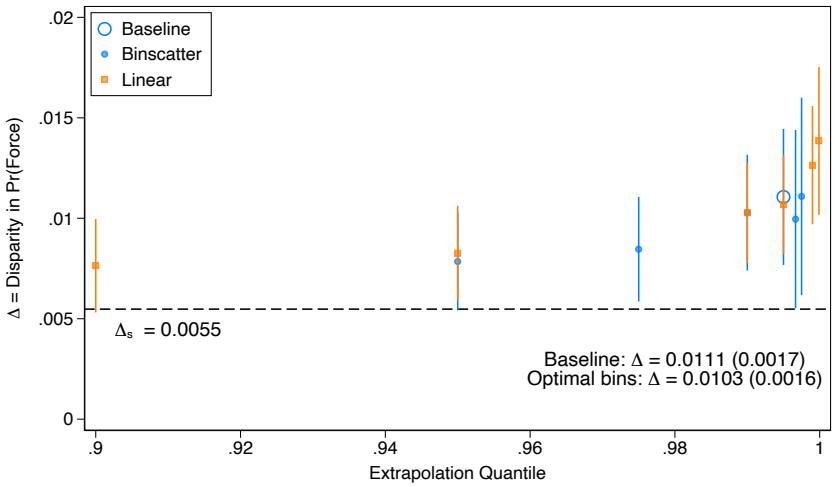
*N*otes: This table reports our baseline estimates of the Black share of the target sample of potential arrestees (π) and the associated selection-adjusted force disparities $\tilde{\Delta}$, computing from the estimated π using equation 4. In both cases, we also report the difference between the unselected 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.

Figure 5: Sensitivity of estimates to extrapolation quantile

(a) Estimated π by extrapolation quantile

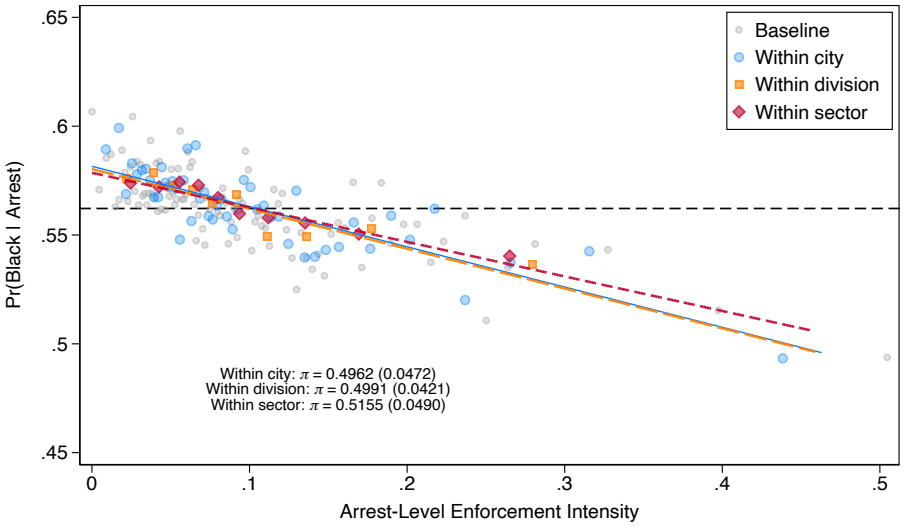


(b) Range of estimated $\tilde{\Delta}$



*N*otes: 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 3: Covariates by race in selected and target samples

	Selected Sample			Target Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	Black	Non-Black	Difference	Black	Non-Black	Difference
Covariate Index	0.0260	0.0245	0.0014	0.0187	0.0185	0.0002
(predicted force)	(0.0022)	(0.0020)	(0.0004)	(0.0017)	(0.0016)	(0.0005)
Female	0.2427	0.2127	0.0301	0.1715	0.1846	-0.0130
	(0.0110)	(0.0043)	(0.0075)	(0.0228)	(0.0180)	(0.0189)
Age	32.8813	33.1241	-0.2428	32.3931	32.8425	-0.4494
	(0.3896)	(0.2041)	(0.2608)	(0.4814)	(0.5667)	(0.5658)
Violent Crime	0.3451	0.2811	0.0641	0.0399	0.0228	0.0170
	(0.0199)	(0.0117)	(0.0110)	(0.0197)	(0.0110)	(0.0099)
Felony Crime	0.5453	0.4752	0.0701	0.0896	0.0594	0.0302
	(0.0340)	(0.0203)	(0.0170)	(0.0307)	(0.0210)	(0.0152)

*N*otes: This table reports the race-specific distribution of covariates in the selected and target samples using the approach explained in section 6.1.

Table 4: Implications of estimated selection patterns for Fryer Jr (2019) estimates

		Overall		Within-X	
	(1) $E(D_i \mid R_i = w, S_i = 1)$	(2) Δ_s	(3) $\tilde{\Delta}$	(4) Δ_s	(5) $\tilde{\Delta}$
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

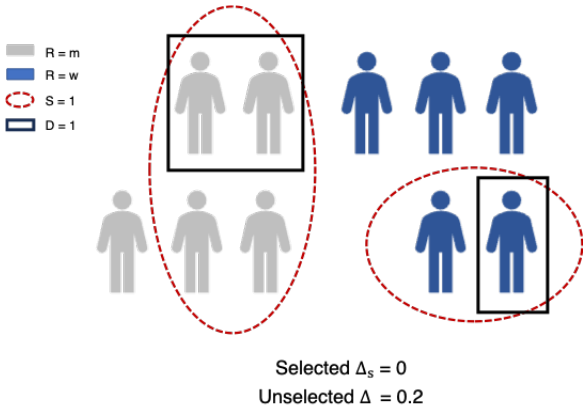
*N*otes: This table reports estimates from Fryer Jr (2019) and selection-adjusted estimates based on the estimated degree of differential selection in our data. Each row corresponds to a different setting and/or specification in Fryer Jr (2019). The first column reports the setting/specification-specific 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

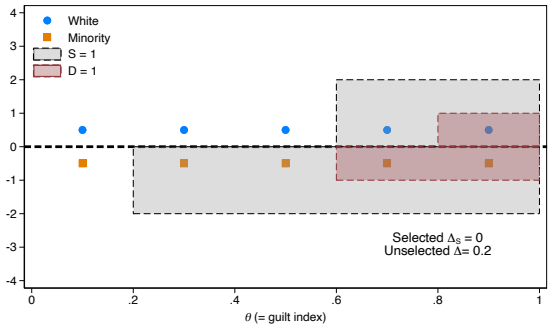
A Appendix figures and tables

Figure A-1: Visualizing the identification problem

(a) Simple version



(b) Version based on Anwar & Fang (2006)



*N*otes: This figure illustrates two approaches for visualizing the identification problem of interest. The first panel presumes no structure on the decision rule for the relevant agent and simply presents ten potentially selected individuals who are selected and then treated. The second panel corresponds to the framework in Anwar & Fang (2006) and assumes that agents rank individuals for selection and treatment based on a "guilt" index denoted by θ. Note that the key moments of interest are identical in each panel. The potentially-selected target sample is comprised of five white and five minority individuals (so $\pi = Pr[R_i = m] = 0.5$). Four minority individuals are selected and two are treated, while two white individuals are selected and one is treated. hence, the selected sample disparity in each case is $\Delta_s = 0$, while the unselected disparity is $\Delta = 0.2$.

Table A-1: 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	134363	75441	58922	104043	30320

*N*otes: This table reports summary statistics for our analysis sample of arrests ($N = 134,363$).

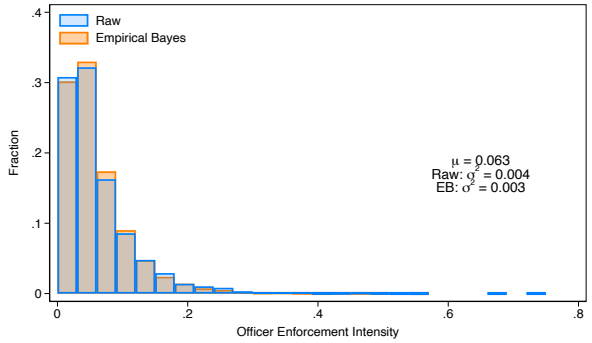
Table A-2: Comparison of Chicago and Seattle to U.S. Cities

	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)

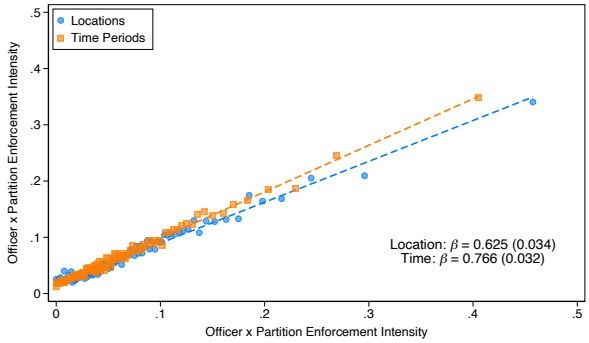
*N*otes: This table combines information from the FBI Uniform Crime Reports (UCR) on felony index crimes and arrests, with fatal use of force records from the crowd-sourced database *Fatal Encounters*, and demographics from the U.S. Census 5-year ACS, between 2014-2018. Arrest, crime, and use of force 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.

Figure A-2: Validating officer-level enforcement intensity

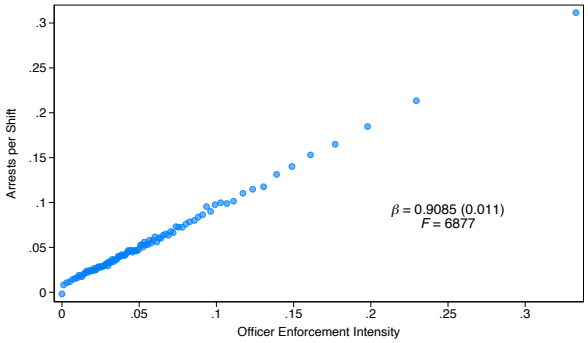
(a) Distribution of estimates



(b) Correlation of estimates



(c) First stage using partitioned instrument



Notes: 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 × 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 on 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 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-2) 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.

Table A-4: Unscaled estimates for target sample

	Target sample estimates:		
	(1) Black	(2) Non-black	(3) 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 estimates 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.

Table A-5: Robustness to specification choices

	Racial Composition			Disparity in Force		
	(1) π_s	(2) π	(3) Difference	(4) Δ_s	(5) $\tilde{\Delta}$	(6) 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)

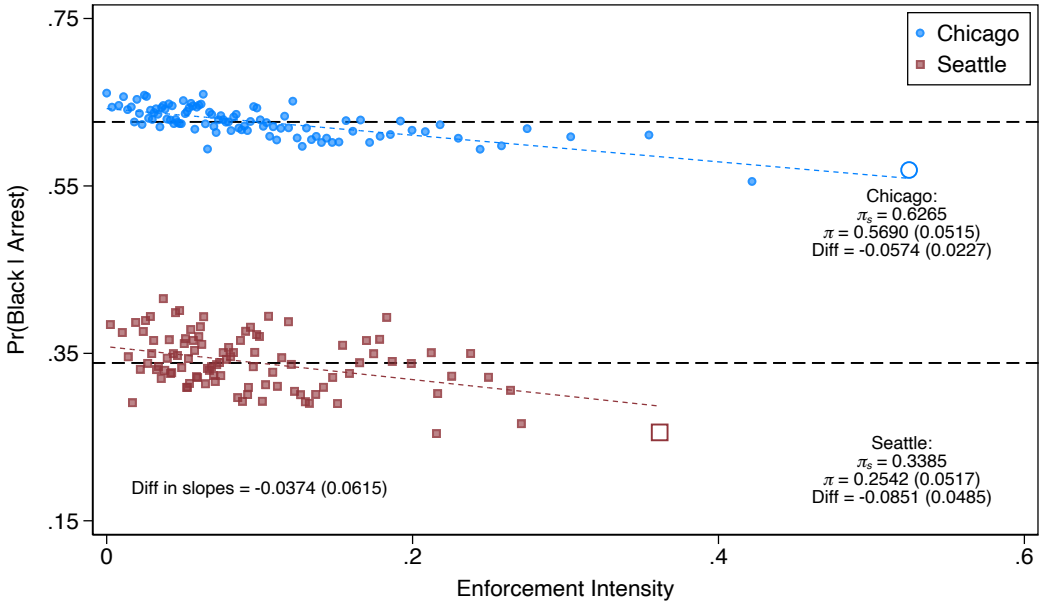
*N*otes: This table presents estimates which are identical to the first row of table 2 except that we vary the measure used to construct enforcement intensity (panel a), the method for aggregating officer-level enforcement intensity into arrest-level enforcement intensity (panel b), and the fixed effects include in the estimation (panel c).

Table A-6: Estimates using empirical bayes shrinkage of officer enforcement intensity

	Racial Composition			Disparity in Force		
	(1)	(2)	(3)	(4)	(5)	(6)
	π_s	π	Difference	Δ_s	$\tilde{\Delta}$	Difference
Combined Binscatter	0.5622	0.5224	0.0398	0.0055	0.0089	-0.0034
		(0.0358)	(0.0139)		(0.0017)	(0.0011)
Combined Linear	0.5622	0.5061	0.0561	0.0055	0.0102	-0.0047
		(0.0428)	(0.0122)		(0.0013)	(0.0008)
Within-city Binscatter	0.5622	0.5163	0.0459	0.0055	0.0094	-0.0039
		(0.0368)	(0.0179)		(0.0020)	(0.0015)
Within-city Linear	0.5622	0.5087	0.0535	0.0055	0.0100	-0.0045
		(0.0424)	(0.0119)		(0.0013)	(0.0008)

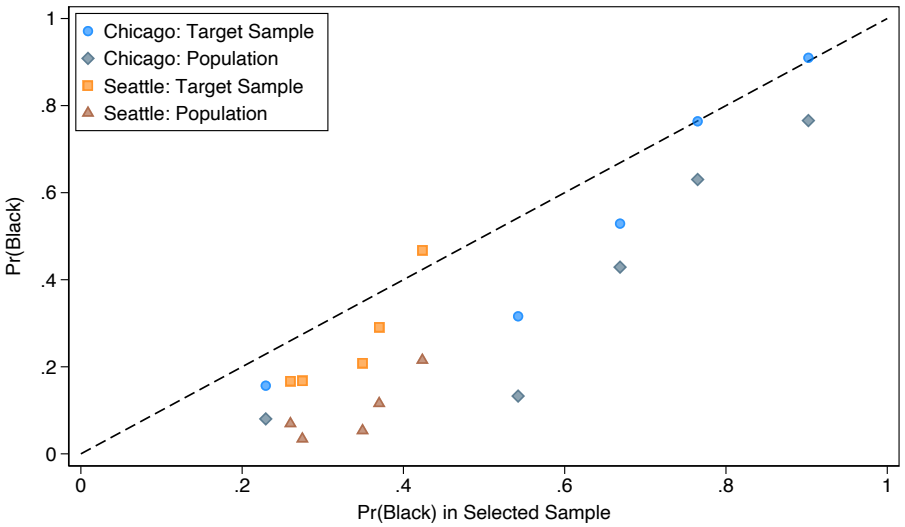
*N*otes: This table presents estimates which are identical to the first row of table 2 except that we apply empirical Bayes shrinkage to the estimated officer-level enforcement intensity measure prior to constructing the arrest-level enforcement intensity measure and estimating π. We report estimates using both a binscatter and linear approach for obtaining the estimated π and approaches which separately shrink the enforcement intensity measures and estimate π by city.

Figure A-3: Setting-specific estimates of π



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

Figure A-4: Division-specific estimates of π



Notes: This figure reports division-specific estimates of the target sample composition. Each point corresponds to a division. The horizontal axis reports the division-specific fraction Black in the selected sample (i.e., π_s^d). Blue circles and orange squares report the division-specific estimates of the fraction Black in the target sample (i.e., π^d) for divisions in Chicago and Seattle, respectively. Blue diamonds and orange triangles report the Black population share for divisions in Chicago and Seattle, respectively. The 45-degree line is depicted with a black dashed line – the vertical distance between the dashed line and the circles/squares indicates the extent to which there is differential selection in each division (i.e., $\pi_s^d - \pi^d$).

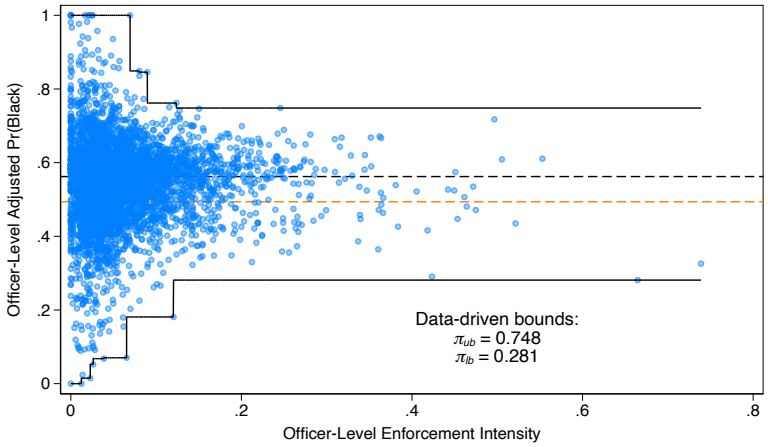
Table A-7: Estimates dropping arrests for officer assault

	Racial Composition			Disparity in Force		
	(1) π_s	(2) π	(3) Difference	(4) Δ_s	(5) $\tilde{\Delta}$	(6) 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)

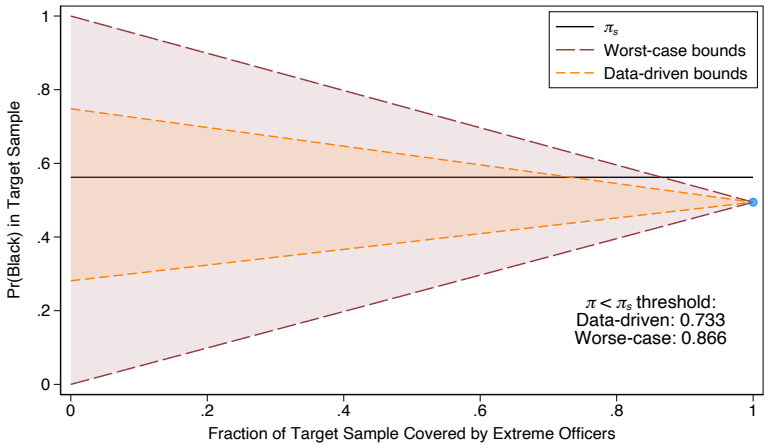
*N*otes: Same as table 2 except that we drop arrests which are for assault on officers ($N = 565$).

Figure A-5: Bounds on π under relaxed assumptions

(a) Constructing data-driven bounds from officer variation



(b) Worst-case and data-driven bounds on π



*N*otes: Panel (a) depicts estimates of the officer-level (FE-adjusted) Black share of arrestees, obtained by estimating to (6) at the officer $\times$ arrest level and weighting by one of the number of officers, and storing the estimated officer effects, against officer-level enforcement intensity. Solid Black lines report the maximum and mininum values among officers with greater enforcement intensity. Dashed black and orange lines depict π_s and π respectively. Note that applying empirical Bayes shrinkage has minimal impact on the data-driven bounds ($\pi_{lb} = 0.321$; $\pi_{ub} = 0.729$). These data-driven bounds on π imply bounds on $\tilde{\Delta}$ of $(-0.0251, 0.0217)$. Panel (b) reports bounds on π as a function of the share of the target sample covered by the extreme officers and under two assumptions about the composition of potential arrestees uncovered by the extreme officer sample.

Table A-8: Moment inequality tests for split-sample monotonicity

	Bootstrap p-value	
	(1) Binscatter	(2) Linear
Black	0.19	0.48
Female	0.04	0.04
Age 0-24	0.11	0.25
Age 25-34	0.13	0.58
Age 35-44	0.73	0.95
Age 45+	0.38	0.33
Violent Crime	0.74	0.93
Felony Crime	0.29	0.33

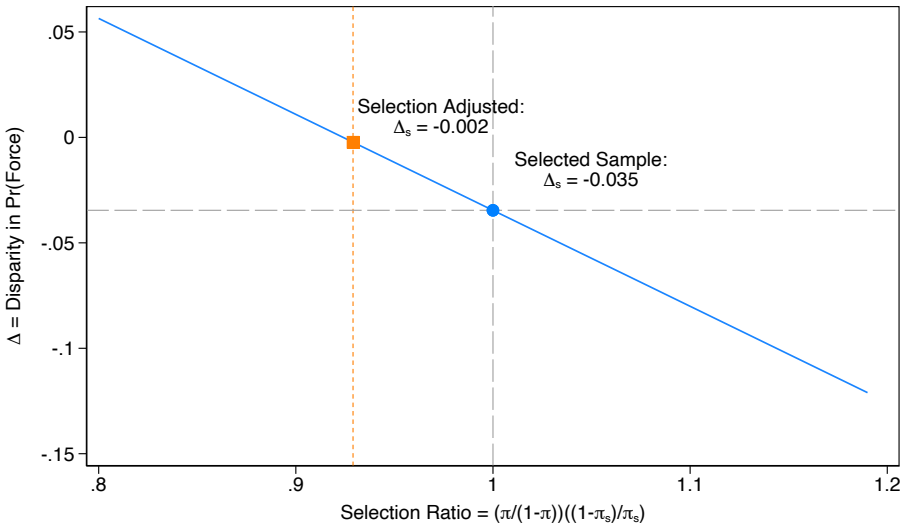
*N*otes: This table reports bootstrapped p-values for the moment inequality test for split-sample monotonicity, described in further detail in appendix C.

Table A-9: 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)

*N*otes: 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 additional 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.

Figure A-6: 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 4.

B 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. In this Appendix, we discuss the legal context for two aspects of this setting: (1) the legal basis for an arrest interaction, and (2) the legal basis for establishing discrimination in police interactions.

B-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.[16] 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, even when charges are later dismissed; a stark example is *District of Columbia vs. Wesby (2018)*,[17] 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). Because the reasonableness of any given officer is subjective, one way to interpret this standard is to ask whether any other officer would have made a particular arrest in a given situation.

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 & Hemmens, 1999).[18] When 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

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

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

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

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.

B-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.[19] 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.[20]

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 suspect demographics and crime characteristics (e.g. Zimroth, 2021). Critically, as noted in our introduction, 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

[19]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

[20]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

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.[21] 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 framework to measure race disparities in use of force, which explicitly addresses this benchmarking problem, and provides a novel approach to selection adjustment.

C Technical Appendix

D 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,

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

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:

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

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

[21] 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

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 minority and white individuals as instruments and construct estimates $\hat{u}_i$ that have independent variation in the outcome equation (B-1). 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.