The Color of Confinement:
Racial Bias and Jail Populations Across America

Lorie A. Fridell

Department of Criminology, University of South Florida, Tampa, FL

Christopher J. Marier

Department of Criminal Justice, University of Central Florida, Orlando, FL

Notes

This is the preproduction version of the accepted manuscript at the American Journal of Criminal Justice. The full, published version can be found at https://doi.org/10.1007/s12103-024-09788-2

Correspondence should be directed to Christopher J. Marier,
christopher.j.marier@gmail.com, 828-278-9591

Lorie A. Fridell is Professor Emeritus of Criminology at the University of South Florida, former Director of Research at the Police Executive Research Forum, and CEO of *Fair and Impartial Policing*, a national law enforcement training program.

Christopher J. Marier is an assistant professor of criminal justice at the University of Central Florida. His areas of interest include race and justice, policing, and the social structure of punishment.

The Color of Confinement:

Racial Bias and Jail Populations Across America

Abstract

This study builds on the body of research examining whether racial disparities in criminal justice can be attributed to bias. The purpose of the current study was to examine whether there is a relationship between aggregate levels of bias and race-specific incarceration rates in U.S. counties. With data from the Vera Institute of Justice, the U.S. Census Bureau, and the Harvard Project Implicit, this study uses county-level estimates of implicit and explicit biases via Multilevel Regression with Poststratification to assess the relationship between those two types of biases and Black and White prisoners in 2,825 county jails across the U.S. using negative binomial regression. Results indicate that pro-White/anti-Black explicit and implicit bias are associated with a higher population-adjusted number of Black prisoners, and fewer White prisoners, even after controlling for socioeconomic covariates and arrest rates. This research provides compelling evidence that racial bias may contribute directly to racial inequity in jail populations and that bias can be understood as a collective phenomenon impacting social systems.

Keywords: explicit bias, implicit bias, prejudice, jail populations, racial disparity, incarceration

Introduction

In the early part of the twentieth century, prisoner population volumes were extraordinarily consistent, prompting some researchers to proffer theories about the "stability of punishment" (Blumstein & Cohen, 1973). But from 1975 through the turn of the century, prison rolls swelled more than fivefold. By 2003, there were 2.1 million people in prison or jail, accounting for nearly one percent of the entire population of the United States. Despite small declines since 2009, the prison boom of the past 50 years is unprecedented in world history. The U.S. incarceration rate currently exceeds that of all other nations at 810 prisoners per 100,000 residents, and 2.5% of the U.S. adult population is currently under some form of correctional supervision (Minton, 2021).

The expansion of incarceration is disconcerting enough, but the impacts were disproportionately borne by Black Americans. In 1880, the Black/White incarceration rate ratio was a bit under 3–to–1. One hundred years later—after decades of so-called "racial progress"—the Black/White incarceration ratio in prisons had more than doubled to 7–to–1. The disparity in incarceration for *local jails* is not nearly as great as that of the prisons but remains stubbornly high at more than 3–to–1 (Zeng & Minton, 2021). The inconsistency between the apparent end of sanctioned governmental racism and actual racial outcomes deserves careful consideration. Why are incarceration disparities so intractable?

While much prior research has established racial disparities in incarceration, much of this research cannot resolve whether these disparities are directly attributable to *bias and discrimination*[1] since most of these studies rely on inferences based upon indirect proxies or assumptions that unexplained variance can be attributed to bias. In this study, we will assess whether racial bias—regarded as a collective phenomenon rooted in social structures—is associated with racial patterns of jail incarceration and whether *implicit* bias is just as harmful as

[1] "Bias" is used here to refer to prejudice against a person. "Discrimination" is used to refer to unjust treatment based on bias.

plainly racist explicit bias. No study to date has examined the relationships between implicit racial biases and the composition of incarcerated populations. The current study uses state-of-the-art methods to produce reliable county-level estimates of implicit and explicit bias and assess the relationship between bias and the incarceration of Black and White prisoners in county jails. A series of count regression models is used to assess the relationships based on data for over 2,800 counties in 45 U.S. states between 2011–2015 using data from the Vera Institute of Justice, the U.S. Census Bureau, and county-level estimates of racial bias derived from Harvard's Project Implicit.

Literature Review

It is not appropriate to assume that the longstanding and substantial racial disparities in jail and prison populations described above are linked only to the decision-making immediately preceding incarceration, such as decisions involving bail and/or sentencing. Disparities that are produced across our criminal justice system could accumulate over the processing of individuals —from the crime-reporting stage through the sentencing stage—to produce the disparity we see reflected in incarcerated populations. Concannon and Na (2024) describe the importance of recognizing the "cumulative disadvantages that accrue across the entire life-course of criminal cases" (p. 3) and aggregate within jurisdictions over time.

Indeed, research has documented disparities at all stages of criminal justice processing. There is disproportionate representation of racial/ethnic minorities as subjects of various police activities such as arrests or tickets (e.g., Durose & Langton, 2013), use of force (e.g., Lautenschlager & Omori, 2018), searches (e.g., Durose & Langton, 2013), and pedestrian or vehicle stops (e.g., Durose & Langton, 2013). Racial/ethnic disparities have also been documented in the workings of prosecutors and defense attorneys (Kutateladze, 2018), judges (U.S. Sentencing Commission, 2018), juries (Levinson, Cai, & Young, 2010), probation and parole professionals (Wilson, 2023), and others.

Racial/ethnic disparities throughout the criminal justice system have been attributed to several factors. One potential contributor to disproportionate representation of Black people in the criminal justice system is differential criminal behavior on the part of their racial group (Hipp, 2011; Loeber, Farrington, Hipwell, Stepp, Pardini, & Ahonen, 2015; Mears, Cochran, & Lindsey, 2016; Sampson & Lauritsen, 1997; Sampson, Morenoff, & Raudenbush, 2005). Some researchers have suggested that the disparity could be explained, at least in part, by facially neutral laws and practices that produce disparate impact, such as the 100-to-1 sentencing disparity for possession of crack versus powder cocaine (Tonry & Melewski, 2008). Notwithstanding evidence for these colorblind explanations for disparity, numerous scholars have also considered the possibility that disparities reflect bias on behalf of citizens and criminal justice personnel—the central focus of the present study.

Researchers have attempted to determine whether and to what extent bias might contribute to the over-representation of Black people in the criminal justice system. Most of these studies have assessed bias *indirectly*, by measuring racial disparities at various points in the system (e.g., police intervention, plea bargaining) and determining how much disparity is left over after the researchers control for the legally relevant factors (e.g., resistance to police, criminal record). This residual disparity is attributed to bias.

A significant drawback to indirect research, however, is the fact that researchers cannot effectively measure all the factors that might legitimately produce disparity. This omitted variable problem can call into question the researchers' conclusions that the residual disparity does in fact reflect bias on the part of actors—individually or collectively. An alternative methodology for isolating bias as a potential cause of disparity is most relevant to the current study. In this "direct assessment" model, researchers measure the level of explicit and/or implicit bias in general populations or on the part of criminal justice actors and then assess whether level of bias corresponds to disparate decision making. While indirect measures of bias estimate the residual disparities after accounting for legally relevant factors, direct measures provide an

understanding of how explicit stereotypes and implicit associations contribute to these disparities. These methods are not mutually exclusive; rather, they provide complementary insights into the mechanisms through which racial bias manifests in incarceration disparities. As a backdrop for discussing direct assessments, the following section examines the difference between explicit and implicit bias.

Explicit and Implicit Bias

Explicit bias is generally what one envisions when thinking about prejudice and bias. With explicit biases, a person associates one or more groups with negative characteristics. These attitudes are based on animus or hostility toward the groups (Amodio & Mendoza, 2010). Explicit biases can affect a person's perceptions and actions, producing discriminatory behavior. It is conscious and deliberate. That is, the person with explicit bias is aware of his/her hostility toward groups and will justify that hostility; s/he is generally unconcerned about the resulting discriminatory behavior (Greenwald & Banaji, 1995; Hardin & Banaji, 2013; Staats, 2013).

Implicit biases are similar to explicit biases in that people link individuals to stereotypes or generalizations associated with their group or groups; these are called "implicit associations" (for reviews see Hardin & Banaji, 2013; Staats, 2013; Staats, Capatosto, Wright, & Jackson, 2016). Unlike explicit biases, implicit biases are not necessarily based on animus or hostility and these implicit associations can impact perceptions and behavior outside of conscious awareness (Devine, 1989; Petty, Fazio, & Brinol, 2012). For instance, many people associate People of Color with crime, aggression and violence (Correll, Hudson, Guillermo, & Ma, 2014; Gilliam, Maupin, Reyes, Accavitti, & Shic, 2016). This implicit association could lead community members or police officers to more closely scrutinize the behavior of People of Color and be more likely to suspect that criminal behavior is afoot. Even individuals who, at the conscious level, reject prejudice and stereotyping can and do harbor implicit biases (Graham & Lowery, 2004; Kang, Bennett, Carbado, & Casey, 2011).

Research that has Measured the Impact of Bias Directly

Researchers attempting to directly assess the impact of human bias on criminal justice disparities have measured explicit bias, implicit bias or both, and this research has used either individual-level measures or aggregate-level measures. For instance, per the former, a researcher might measure the level of bias in individual prosecutors and then determine whether that bias is linked to disparity in their decisions, such as in responses to written scenarios (see e.g., Rachlinski, Johnson, Wistrich & Guthrie, 2009; Sim, Correll & Sadler, 2013).

Several research teams have argued, however, that aggregate-level studies of bias are superior to individual-level studies because an individual's personal level of bias is not the only factor that might produce discriminatory decision making (e.g., Ekstrom, Le Forestier, & Lai, 2022; Hehman, Flake, & Calanchini, 2018). Jurisdiction-level biases capture the culture that prevails in the criminal justice system and in the community as a whole. Even an individual without racial bias may act in a biased fashion based on the expectations of colleagues, superiors, and community members. Agency leaders may themselves have high levels of bias and/or they may be acting in accordance with the expectations of the community members whom they serve. Aggregate levels of bias capture these collective racial attitudes. Hehman et al. (2018), in their study of disparities in police use of lethal force, make an additional argument —pointing to bias "contagion" within geographic areas. They write that, "prevailing regional biases might shape police officers' own attitudes, and their behaviors on the job are the result of these attitudes" (p. 394).

Various researchers have examined bias in the aggregate and then assessed whether collective bias in a particular geographic location is associated with predicted societal outcomes, including disparities. Research has demonstrated that geographically-aggregated bias measures from the IAT predict expected outcomes such as participation in demonstrations/rallies (Zerhouni, Rougier, & Muller, 2016), infant health (Orchard & Price, 2017), school discipline (Riddle & Sinclair, 2019), health care availability (Leitner, Hehman,

Ayduk, & Mendoza-Denton, 2016a), Medicaid expenditures (Leitner et al., 2018), racial disparities in health outcomes (Leitner, Hehman, & Snowden, 2016a, 2016b; Miller, Varni, Solomon, DeSarno, & Bunn, 2016; Orchard & Price, 2017), and gender disparities in math and science achievement (Nosek, Smyth, Sriram, Lindner, Devos, Ayala,…Greenwald, 2009). This research is made possible by the fact that Project Implicit has aggregated IAT scores within geographic areas. Pinkston (2015) has shown that IAT respondents show patterns of bias that are similar to that of nationally representative samples.

Two studies have used the aggregate-bias methodology to predict criminal justice disparities, specifically in both cases, law enforcement disparities (Ekstrom et al., Hehman et al., 2018). As previewed above, Hehman et al., (2018), for instance, examined collective bias and racial disparity in lethal use of force. The researchers used aggregate IAT data for both explicit and implicit bias and produced separate measures of each for White people and for Black people within Core-Based Statistical Areas (CBSAs). Hehman et al., (2018) found that the aggregate level of *White* residents' *implicit* bias predicted disproportionate lethal force against Black people. There was no relationship between *Black* residents' implicit biases and lethal force disparities. For both Black and White residents, there was no relationship between *explicit bias* and lethal force disparities.

The Current Study

The current study uses advanced statistical methods to examine the relationship between county-level variations in racial bias and the incarceration of both Black and White individuals in local jails. We combined jail incarceration data from the Vera Institute of Justice with racial-attitude data from Harvard's Project Implicit dataset, plus socioeconomic controls from the U.S. Census. We hypothesized that pro-White/anti-Black bias would be associated with higher Black incarceration and lower White incarceration.

Methods

The data come from the Vera Institute of Justice (2018), Harvard's Project Implicit (Xu, Nosek, Greenwald, Ratliff, Bar-Anan, …Axt, 2019), and the U.S. Census Bureau (n.d.). These data span the years 2011–2015. More detailed information about the data used from each source is provided in the Measures section below.

Sample

The sampling frame for this study consisted of the population of U.S. counties. After listwise deletion due to missing data,[2] the final regression sample consisted of 2,825 U.S. counties. Several states were dropped due to missing data; these states operate combined jail/prison systems and/or do not provide disaggregated statistics about local jail populations under examination here. The following states are omitted: AK, CT, DE, HI, RI, and VT. The resulting 2,825 counties include the vast majority (>96%) of all U.S. residents. The data were clustered within the 45 remaining states and the District of Columbia. On average, there were 64.2 counties observed within each state (ranging from 10 to 223).

Summary statistics appear in Table 1. Black people had lower employment rates, lower marriage rates, and higher poverty than their White counterparts, demonstrating their socio-structural disadvantage. On average, each county had a Black incarceration rate of about 2,960 and a White incarceration rate of about 394 (counts, not rates, are reported in the tables).[3] These figures are right-skewed; the median Black incarceration rate is 1,082, and the median White incarceration rate is 265. Figures 1 and 2 demonstrate the geographic distribution of incarceration by race, as well as the substantial contrast between the incarceration of Black people and White people throughout the U.S.

[TABLE 1]

[FIGURE 1]

[2] The most common reasons for missingness are an absence of poverty data (n = 171), bias estimates (n = 125), employment data (n = 124), and/or jail data (n = 121).

[3] These rates are per 100,000 Black or White county residents (respectively) over the age of 15. See Measures section.

[FIGURE 2]

Measures

This study incorporates two dependent variables, two independent variables, and ten control variables. Each type of variable is described below.

Dependent Variables: Black and White Prisoners in County Jails

Two dependent variables are explored in this study: the number of Black prisoners in county jails, and the number of White prisoners in county jails. These numbers are single-day counts taken each year, summed to represent the total number of Black or White prisoners over the study period. The data come from Vera Institute of Justice (2018), which combines data from the Bureau of Justice Statistics (BJS) Census of Jails (which is more comprehensive) with data from the BJS Annual Survey of Jails (which is more frequent) to produce the incarceration trends dataset that was used in this study. The data set provides information on inmates held beyond arraignment, whether pre-trial or post-conviction, and therefore captures the cumulative effect of a range of potentially racially-influenced criminal justice processes, including bail determinations, conviction, the in/out decision, and sentence length.

We focus on county jails rather than state or federal prisons for several reasons. First, studying counties rather than states permits a much larger sample and many more degrees of freedom for analyses. Furthermore, state-level analysis is less compelling when considering the substantial variation in bias and community characteristics within states (across counties); consider, for instance, that the Florida or Texas panhandles are quite different from their urban centers, differences which state-level averages obfuscate. The determination of who gets locked up is primarily determined by county-level officials: police, judges, prosecutors, county jail administrators, and so on. The determination of who gets investigated, arrested, and jailed is also determined by county-level community members: complainants, victims, witnesses, and jurors. Jail data capture a greater range of lower-level crimes where discretion is greater and

bias is more likely to manifest (Dovidio & Gaertner, 2000; Hester & Hartman, 2017).[4] They also include pre-trial detainees, reflecting disparities in pre-trial detention that may be driven by regional biases. All of this is to say that a more granular, county-level examination is warranted, given that a great deal of the criminal justice process—and the opportunity for insidious bias—occurs at the local level.

Independent Variables: Implicit Bias and Explicit Bias

The focal independent variables in this study consist of county-level estimates of implicit and explicit bias. This choice is rooted in social psychologists' view that aggregate levels of bias reflect *concept accessibility*—a "wisdom of crowds" effect–that recognizes implicit bias as socially constructed and context-dependent rather than a stable individual trait (Payne *et al*., 2017). This is akin to concepts such as "organizational culture" and "social capital" that are measured with surveys of individuals but are understood to represent collective, higher-order phenomena. For instance, Sampson, Raudenbush, and Earls' (1997) seminal *Science* article studying collective efficacy notes that their measures "were aggregated to the neighborhood level" (p. 920). Sociologists and political scientists have a long history of aggregating individual attitudes into meaningful collective constructs (Esmer & Patterson, 2007).[5]

The measure of implicit bias comes from the Implicit Association Test (IAT) produced by the Harvard-based Project Implicit (PI). The IAT, based on response latency, is the most widely used method for measuring implicit bias (Xu et al., 2019). Staats explains (2013, p. 24): "These measures rely on reaction times to specific tasks in order to uncover individual's biases … The underlying premise of these reaction-time studies is that individuals are able to complete cognitively simple tasks relatively more quickly than those that are mentally challenging." In taking the IAT, respondents are timed as they sort concepts. Since people are faster at

[4] Furthermore, prisoners are frequently held in state prisons outside their original county of residence or offense, a fact that confounds county-level social control responses.

[5] It is therefore of some surprise that the aggregation of biases has perhaps drawn more controversy than any other individual attitude (Payne et al., 2017; Jost, 2019).

completing simple tasks than challenging ones, they will be faster "sorting" concepts that are *linked* in their heads than sorting concepts that are *not linked*. As such, if a person is faster at linking "women" and "childcare" than "men" and "childcare," it can be inferred that the person holds a cognitive association in his/her head between women and childcare (but not men and childcare).

The present study uses data from the "Race IAT" available through the Project Implicit Demo Website Dataset (https://osf.io/y9hiq/). The Race IAT measures the extent to which respondents associate Black and White faces with either "good" words (e.g., joy, wonderful, pleasure) or "bad" words (e.g., evil, terrible, awful). Higher scores reflect the extent to which "good" is conceptually associated with White people and "bad" is conceptually associated with Black people—suggesting increased implicit bias toward Black people. The measure of explicit bias reflects the "thermometer scale," which is one of the most commonly used "global measures" of racial prejudice and that also comes from Project Implicit. Respondents use 11-point scales to respond to two questions: "How warm or cold do you feel toward African Americans/Black people" and "How warm or cold do you feel toward European Americans/White people." The difference between the two scores represents explicit bias, where (like the IAT score) higher values represent more pro-White/anti-Black bias. Both the implicit bias and explicit bias scores are geographically identified so that they can be aggregated to units such as counties.

We addressed a major limitation with producing county-level estimates from the IAT data: the Project Implicit sample is non-random. Participants self-select to take the IAT. Therefore, we used county-level measures that have been estimated via Multilevel Regression with Poststratification (MRP) by Ekstrom, Lai, & Le Forestier (2019) and Ekstrom, et al. (2022) because of its distinct advantages for accurately estimating public attitudes based on nonrandom samples. Surveys often struggle to get a representative sample of the population.

MRP helps to correct these biases through a two-stage process involving hierarchical linear modeling followed by a poststratification process that applies weights from the known population characteristics derived from census benchmarks. As Ekstrom et al., (2022) explain, multilevel regression with poststratification "first estimates the *direction and intensity* of residents' racial attitudes within demographic subgroups in each county. It then uses the *prevalence* of each subgroup within the county population to weight its contribution to a county-level attitude estimate" (p. 499). U.S. Census county-level estimates of race (White, Black, other), gender (male/female), and age (an ordinal measure with eight categories) were used to produce MRP estimates (Ekstrom et al., 2022). MRP can generate estimates for small subgroups or small geographic areas, even when the sample size for these groups is very small (Zhang *et al*., 2014). This is possible because the multilevel model borrows strength from the entire sample when making estimates for individual groups, and the poststratification ensures these estimates are representative of the known population characteristics. MRP performs well, producing stable estimates that outperform alternative approaches (Downes & Carlin, 2020), even with highly non-representative samples (Wang, Rothschild, Goel, & Gelman, 2015). The use of MRP estimates is well-established in peer-reviewed research exploring the effects of implicit bias on various public health outcomes (Eckstrom *et al.*, 2022; Riddle & Sinclair, 2019; Orchard & Price, 2017; Leitner *et al*., 2016a, b).

At the individual level, IAT scores range from -2 to +2 and explicit bias scores range from -10 to +10, where scores of 0 indicate no Black/White preference and positive scores indicate pro-White/anti-Black bias. As illustrated in Table 1, county-level implicit bias estimates ranged from -.024 to .417, with an average of .338, indicating moderate anti-Black bias. County-level explicit bias estimates ranged from -1.542 to +1.078, with an average score of .443, indicating moderate anti-Black bias. These scores were standardized in the regression models for two reasons. First, they lack an inherent intuitive metric, so it is useful to speak about relationships in terms of standard deviations (e.g., "a 1-standard deviation increase in implicit bias was

associated with a X% change in the Black incarceration rate ratio"). Furthermore, standardization makes comparisons between implicit bias effects and explicit bias effects somewhat more comparable. Standardization means that a value of zero reflects the "average" bias across counties, rather than no bias, but does not change the underlying relationships or statistical inferences.

Control Variables

Given that criminal justice outcomes may reflect the influence of other systemic disparities and inequities, many of which are correlated with both bias and criminal justice involvement, we introduce several control variables. Control variables come from 2011–2015 American Community Survey (ACS) 5-year estimates from the U.S. Census Bureau (specifically, tables S1201, S1701, and S2301) and from Kaplan's (2020) concatenated UCR data. These variables include both the Black and White employment/population ratio (the percentage of each group employed, whether looking for work or not), both the Black and White marriage rate (the percentage of each group over the age of 15 that are married), and both the Black and White poverty rate (the percentage of each group who fall beneath the official poverty line). The Black and White arrest rate for all crimes (the number of arrests per 1,000 residents) is included to control for race-specific offending (Kaplan, 2020).[6] Finally, as described below, the race-specific county populations between the ages of 15 and 64 are included as an exposure term in regression models.

Analytic Strategy

We employ a series of negative binomial regression models to predict the number of Black and White prisoners in each county jail. Count models, such as negative binomial regression, are appropriate when the dependent variable is a discrete count (such as the

[6] The use of race-specific arrest rates as proxies for racially differential offending is obviously problematic since patterns of arrest may be confounded with the biases of community members and police officers who invoke the law. Thus, this is perhaps an overcorrection. That we still find evidence of bias in incarceration after controlling for differential arrest rates speaks to the strength of the relationships between bias and incarceration.

number of prisoners) and offer several advantages over ordinary least squares (OLS) regression (Hilbe, 2011; 2014). For instance, prisoner counts (and rates) are highly right-skewed—most counties have few prisoners, and a few counties have many prisoners. Typically, this violates assumptions that the errors are normally distributed and have constant variance. Sometimes this is overcome by log-transforming incarceration rates, but this creates several problems of its own; counts of "0" cannot be log-transformed, for instance, and so they are either excluded from the model or they are added to an arbitrary constant before log transformation. While log-transforming the dependent variable in OLS models can be a simple and quick approach, it does not fully address the specific characteristics and assumptions of count data. Count models offer a more appropriate and comprehensive framework for analyzing count outcomes, leading to better model fit, interpretation, and inference (Cameron & Trivedi, 2013).

The negative binomial regression models predict the number of prisoners, but they actually estimate the *incidence rate* of incarceration by using an exposure term with a coefficient fixed to 1. When predicting the number of Black prisoners, that exposure term is the number of Black county residents who are between the ages of 15 to 64 over the 5-year study period (and, likewise, White residents between 15 to 64 when estimating White prisoners). Therefore, these models account for each county's Black/White population when estimating Black/White prisoner counts.

Like the prisoner count, the exposure term encompasses a 5-year count (person-years). Thus, the models predict race-specific incarceration rates—the probability that a Black or White resident between the ages of 15 and 64 was incarcerated in a local jail between the years 2011 and 2015.

Negative binomial regression was chosen over Poisson count models because there was evidence of overdispersion of prisoner counts, and model fit statistics (including LL, AIC, and BIC) were significantly better with negative binomial models. Negative binomial regression

uses an extra dispersion parameter, *alpha*, to model the increased variation in the data, rather than assuming that the conditional variance approximates the conditional mean, as Poisson models do (Hilbe, 2011; 2014).

There were two peculiarities of the data: excess zeros (zero-inflation) and multilevel clustering of counties within states. However, these two issues could not be statistically modeled simultaneously; to our knowledge no statistical packages can simultaneously estimate zero-inflated models and multilevel models. We therefore report, first, a series of zero-inflated models that do not account for the multilevel structure, and second, a series of multilevel models without adjustment for excess zeros. In the zero-inflated models, clustered standard errors were used to account for heteroscedasticity and correlation within states (Angrist & Pischke, 2008). In zero-inflated models, we model excess zeros as a function of crime (counties with little to no crime are more likely to have no prisoners) and total population (counties with little to no residents and little to no tax base are more likely to have no prisoners).

We report one set of analytic choices here, but in fact explored several alternative model specifications. For instance, we specified more parsimonious negative binomial models without zero-inflation or multilevel modeling; models using traditional, multilevel, and zero-inflated Poisson regression; as well as models using state fixed effects to more strictly control for state-level heterogeneity. We also specified alternative models limited to counties that had at least 50 IAT respondents and populations that were at least 2% Black. Ultimately, each of these models converged toward the same substantive inferences,[7] indicating that the results are robust to a wide variety of models and specifications.[8] (See the online supplement.)

[7] In every supplemental model specification, increases in county-level anti-Black implicit bias and explicit bias were associated with a higher number of population-adjusted Black prisoners, consistent with the models reported here. However, several specifications diverge from the main models for White prisoners, where they generally show non-significant relationships between bias and White prisoner counts. Nonetheless, model fit is generally poorer in the supplemental models.

[8] It's worth reiterating that our "sample" of counties is much closer to a population than a sample. Furthermore, even if one were to assume that the set of counties studied herein were a sample, it is not a random sample. For these reasons, it is necessarily true that statistical

Results

A series of zero-inflated negative binomial regressions was used to model Black and White prisoner populations in county jails, as predicted by county-level MRP estimates of implicit bias and explicit bias net of controls (Table 2), after accounting for excess zeros in counties with little crime or few residents.[9] The coefficients have been exponentiated to incident rate ratios (IRR's), which here reflect incarceration rate ratios. In short, both forms of anti-Black bias were significant and substantive predictors of population-adjusted race-specific prisoner counts in county jails. A one standard deviation increase in implicit bias was associated with a 26.1% increase in the Black incarceration rate ratio ($p < .001$) and a 20.8% decrease in the White incarceration rate ratio ($p < .001$). A one standard deviation increase in explicit bias was associated with an 24.1% increase in the Black incarceration rate ratio ($p < .001$) and a 22.8% decrease in the White incarceration rate ratio ($p < .001$).

[TABLE 2]

The relationships between bias and incarceration re-emerged in multilevel negative binomial regression models that considered the clustering of counties within states (Table 3), since state law and policy are likely to systematically affect patterns of jail confinement. In these models, a one standard deviation increase in implicit bias was associated with a 23.6% increase in the Black incarceration rate ratio ($p < .001$) and a 14.9% decrease in the White incarceration rate ratio ($p < .001$). Similarly, a one standard deviation increase in explicit bias was associated with a 22.4% increase in the Black incarceration rate ratio and a 14.6% decrease in White incarceration rate ratio.

[TABLE 3]

---

significance and inference are not appropriate considerations, and that the size and direction of the parameter estimates are far more relevant than significance levels, and readers should interpret accordingly.

[9] Implicit bias and explicit bias are modeled in separate models, since VIFs indicated problematic multicollinearity when modeled simultaneously.

As shown in Tables 2 and 3, several control variables also had significant associations with incarceration. White and Black poverty rates predicted increases in both Black and White jail incarceration. Black employment significantly predicted lower Black incarceration and White employment predicted lower White incarceration. Perhaps as a reflection of overall more conservative values, places with higher White marriage rates demonstrated significantly higher incarceration for both races; Black marriage rates, on the other hand, were not associated with incarceration rates. Black arrest rates predicted higher incarceration rates for both Black people and White people in multilevel models (Models 5-7), but no effects for either group in zero-inflated models (Models 1-4). White arrest rates were associated with higher White incarceration rates in every model.

Overall, the results across the models in Tables 2 and 3 (and several other supplemental analyses available in the online supplement) demonstrate that both implicit and explicit biases are associated with county jail incarceration. Specifically, pro-White/anti-Black bias is associated with a higher population-adjusted number of Black prisoners, even when several social structure covariates, including employment, poverty, and arrests, are taken into account.

Discussion

Racial disparities in jail populations remain substantial despite decades of supposed racial progress. The three-to-one ratio of Black prisoners to White prisoners could be due to differential involvement in criminal behavior or other legally relevant factors. It is also possible that implicit or explicit bias on the part of community members or criminal justice officials contributes to this disparity. The purpose of the current study was to examine whether there was a relationship between collective bias and race-specific incarceration rates in local jails using data for over 2,800 U.S. counties, after controlling for differential offending and socio-structural disparities.

The results demonstrate a strong relationship between pro-White/anti-Black bias and levels of Black incarceration rate ratios net of controls. This relationship was found for both

implicit and explicit bias. Both zero-inflated and multilevel negative binomial regression models found that a one standard deviation increase in implicit or explicit bias is associated with a 22 to 26% higher incarceration rate ratio for Black people, and a 15 to 23% lower incarceration rate ratio for White people. These numbers indicate that regional bias is not only statistically significant, but *substantial*, when it comes to the number of Black and White prisoners held in jails. This research joins a growing body of evidence suggesting that collective bias contributes to racial inequities in various social systems, including health systems, education systems and, as here, criminal justice systems (e.g., Orchard & Price, 2017; Riddle & Sinclair, 2019; Hehman et al., 2018).

What manifestations of *implicit bias* on the part of individual criminal justice professionals could produce the substantial increase in Black incarceration rates? There may be various implicit associations between Black individuals and stereotypes that produce the harsher criminal justice treatment, but certainly the particular implicit association between Black individuals and crime, violence, and aggression is a likely explanation. As stated previously, this particular association might lead community members or police officers to focus more attention on the behaviors of Black citizens. They might interpret the same behavior on the part of Black and White people as more suspicious when engaged in by Black people. A prosecutor who associates Black people with a greater likelihood of crime and/or violence might behave more harshly toward Black defendants—in the context of charging decisions, bail recommendations, plea deals, or sentencing recommendations. A judge's implicit biases could affect his/her bail decision or sentences. Juries may be more likely to convict Black defendants. As explained above, race might affect these criminal justice actors absent their intention to discriminate and, indeed, outside of their conscious awareness.

Our results suggest that explicit bias demonstrates similarly large effect sizes as implicit bias. Explicit bias could lead to discriminatory decisions on the part of individuals similar to those described above, but the harm could be even greater than that produced by implicit bias.

Explicit bias is more overt; it is conscious and the person with explicit bias is unconcerned about discriminatory behavior. Because it is more visible, this form of bias might be more effective at producing a biased culture that affects criminal justice decision makers. Such a culture might elect city council members or appoint a city manager who choose *not* to hire the police chief candidate who espouses diversity in hiring and developing trust within marginalized communities. The jurisdiction prosecutor might be criticized for diverting juvenile offenders to treatment or adopting in-house practices designed to reduce the risk of biases in case processing. Similarly, the person facing a judicial election (or reappointment) will not be campaigning in that culture on the fact that he and the rest of his staff will engage in implicit bias training.

Implicit and explicit bias are very highly correlated across counties, consistent with arguments that bias is dependent upon context, systems, and social structures. It is possible that prevailing regional explicit biases and systemic biases produce implicit biases in others living in those contexts, even those without intent to discriminate. Thus, places with higher levels of explicit bias (or more people who hold these biases) may also produce higher levels of implicit bias via contagion (Hehman et al., 2018), concept accessibility (Payne et al., 2017), organizational cultures (Camp, 2023; Hetey, Hamedani, Markus, & Eberhardt, 2024), and social structures (Skinner-Dorkenoo et al., 2023).

Policy Implications

Insofar as the results from this study indicate that collective bias has a substantial (and troublesome) association with jail disparities, interventions likely must be targeted to both individuals *and* aggregates (communities, organizations, and social systems). A key intervention is providing implicit bias awareness training to personnel, the purposes of which are to raise individuals' awareness of implicit biases, identify the negative consequences of biased behavior, and to provide individuals with skills for reducing and managing biases. Implicit bias training programs have been developed for various criminal justice actors such as law enforcement

(Fridell, 2017; Fridell & Brown, 2015), prosecutors and defense attorneys (American Bar Association, n.d.), judges (Casey, Warren, Cheesman, & Elek, 2013), and jurors (Gayla, 2017).

But training for criminal justice system actors may not be enough. Community members exercise the law—by deciding which people and behaviors warrant a call to the police, by choosing whether or not to cooperate or press charges, by choosing whether crime control is conducted via policing versus community organizing, and by voting for candidates and policies that could lead to disparate impacts. Each of these choices, among many others, may be influenced by implicit or explicit biases. Each of these choices also turns individual biases into *systemic biases* that reinforce the implicit cognitive associations that create and sustain biases within individuals, in a mutually reinforcing process. This suggests, perhaps, that understanding and managing implicit social cognition is an essential component of civic responsibility that must be taught to *all* citizens, as other civics topics are.

Smith and Levinson (2012) suggest that hiding demographics from criminal justice decision makers can reduce the risk of bias. “Category masking,” they explain, is the process of hiding the demographics (e.g., in the file) of the person about whom the criminal justice actor is making a decision. (See also Sah et al., 2015.) This can’t be applied within all criminal justice components or at all stages but is particularly applicable to the early case processing of prosecutors. In fact, in 2019, the District Attorney in San Francisco adopted a policy and the needed technology to remove racial information from the police reports that the attorneys used to make initial charging decisions (Sernoffsky, 2019).

Because greater discretion in decision-making is linked to the greater potential for bias (Skeem, Montoya & Lowenkamp, 2023), biased behavior can be reduced if the discretion associated with high-risk decisions is limited/structured (Payne & Vuletich, 2018; Richardson & Goff, 2013; Skeem et al., 2023). It is arguably impossible and undesirable to rid criminal justice decisions of all subjectivity, but criminal justice units (e.g., police departments, prosecutors’ offices) can develop or enhance standards or criteria for decision making. Sentencing formulas

limit the discretion of judges. Prosecutor's offices can similarly adopt clear criteria for various decisions such as charging or bail. Because requesting consent to search allows for considerable police discretion, some chiefs have added standards such as requiring written consent of the subject and supervisor approval prior to conducting a consent search (Nicklas, 2012).

Measuring levels of disparity associated with an individual's decision points (e.g., bail recommendations, sentencing) and providing feedback to that decision maker can serve to prevent future biased decisions in individuals who are motivated to be impartial. Richardson and Goff (2013), for instance, suggest that such data be collected for public defenders and that supervisors require defenders to explain racial disparities appearing in their case load. Smith and Levinson (2012) recommend this for prosecutors. Wistrich and Rachlinski (2017) propose the auditing of individual judges' decisions including bail setting and sentencing. Judges might be compared to their peers across measures of disparity in decision making and the judges informed if they are on the "high end" of racial disparity.

Efforts to reduce the risk of bias in policing have included reducing the volume of activities with high risk for bias (Fridell, 2017). For instance, several cities (e.g., Duluth, San Francisco, Memphis) have adopted policies to reduce the volume of traffic stops for low-level offenses (Lawler, 2023; Pena, 2024; Valencia, 2023). In Philadelphia, police were directed to reduce enforcement of quality-of-life violations (Griffin, 2024). Los Angeles County reduced bike stops (Colligan, 2022); New York City reduced the volume of stop and frisk (White & Fradella, 2016).

Racially diverse criminal justice agencies could reduce the risk of biased behavior, though not because under-represented groups are bias-free; on the contrary even under-represented groups have biases (Kang & Banaji, 2006) and they can even have biases against their own groups (Zeng et al., 2023). Diversity can be helpful because the interaction of diverse individuals within a unit can reduce individual biases as explained by the Intergroup Contact

Theory. Considerable research has shown that positive contact with people who are different from you can reduce both explicit and implicit biases (for a meta-analysis, see Pettigrew & Tropp, 2006). In promoting diversity within prosecutors' offices, Smith and Levinson (2012, p. 826) suggest, "we would expect that a diverse prosecutor's office might not only facilitate an atmosphere with less implicit bias but could perhaps also lead to even more thoughtful and efficient decision-making."

The Intergroup Contact Theory has application at the community level, too, highlighting the importance of policies aimed at promoting racial integration in residential, educational, and workplace settings. Segregation in neighborhoods, workplaces, and classrooms in the U.S. perpetuates individual-level racial biases (Skinner-Dorkenoo et al., 2023). These forms of segregation are largely due to historical policies like redlining, which graded neighborhoods by perceived financial risk; they disproportionately affected areas with populations of color, and led to long-term wealth disparities and reinforced stereotypes that associate White people with wealth and success. Policies to promote integration could involve revising zoning laws, enhancing fair housing enforcement, incentivizing diverse communities, and restructuring school districting. While larger numbers of "out-group" members tend to increase bias for "in-group" members, this tends to be offset in places where intergroup contact is both positive and frequent (Rae, Skinner-Dorkenoo … & Hewstone, 2022). Additionally, policies could focus on fostering intergroup contact by encouraging inclusive community events, diversity training, and cooperative activities across different racial groups. Stringent anti-discrimination enforcement in housing valuations and investments in infrastructure and resources for these communities could address the devaluation of neighborhoods predominantly inhabited by people of color. By confronting these systemic issues, policymakers can work towards dismantling the legacy of segregation and reducing the racial biases it perpetuates.

Ultimately, the types of interventions necessary to overcome the influence of regional racial bias on biased patterns of incarceration are likely to require systemic changes to deep-

rooted social structures (Skinner-Dorkenoo et al., 2023). Patterns of power and privilege in the United States set a foundation for racial biases, influencing individual perceptions and societal structures. Many Americans grow up seeing predominantly White individuals in respected roles, which can lead them to perceive that the country primarily “belongs” to White people. This environment fosters a mindset, especially in children, which endorses existing social hierarchies and resource disparities, often favoring those already more fortunate. Such biases are not only internalized in childhood but are also reflected in adult behaviors, where structural racial inequalities contribute to biases against people of color. Individual-level racial biases contribute to systemic racial inequalities, affecting various domains such as voting behavior, public health responses, and educational disparities. These biases reinforce the systemic inequalities that created them, perpetuating a cycle where beliefs in inequality justify the structures of inequality themselves, thus maintaining the status quo of racial disparities. A radical restructuring of power and wealth is likely necessary to produce greater parity in social institutions, including populations in county jails (Skinner-Dorkenoo et al., 2023).

Strengths, Limitations and Research Implications

This study makes both theoretical and empirical contributions to the study of racial disparities in the criminal justice system. First, it provides evidence that racial bias predicts racial inequity in jails, without the use of indirect proxies for bias or assumptions that unexplained variance can be attributed to bias. By including direct measures of bias and legally-relevant criteria, this study provides evidence that *both* differential offending *and* differential response likely contribute to racial disparities in incarceration. While this study provides evidence that “residual disparity” observed in prior research may, indeed, be partly due to racial bias, much unexplained variance remains, suggesting that its mechanisms are still incompletely understood. Second, it provides evidence that bias can be understood as a collective phenomenon impacting social systems, not merely an individual phenomenon that impacts individual cases. It does so using advanced statistical techniques applied to a sample of nearly

all U.S. counties, robust to a variety of model specifications. Third, the study demonstrates that *both* implicit and explicit bias strongly predict Black and White prison rolls. Those contributions notwithstanding, there are a number of weaknesses of the current study with implications for future research.

Implicit bias is measured in this study using aggregated IAT data. The use of these data is a strength or a weakness depending on where one stands with regard to the ongoing academic debate on the validity, reliability, and generalizability of the IAT. (For summaries of the debates, see Mitchell & Tetlock, 2017; Nagai, n.d.; Oude Maatman, 2017; Mitchell & Tetlock, 2017; Oude Maatman, 2017).[10] There are some studies that report low reliability of the IAT, such as test-re-test reliability (Bar-Anan & Nosek, 2014; Rezaei, 2011). In contrast are the studies that show evidence of both internal consistency and test-retest reliability (Cunningham, Preacher, & Banaji, 2001; Nosek, Greenwald, & Banaji, 2007). (For reviews of test-retest studies, see Gawronski, Morrison, Phills, & Galdi, 2017; Lane et al., 2007.)[11] Jost (2019, p. 2) claims, “The IAT exhibits higher (within-persons) test-retest reliability than other response-latency measures commonly used in psychological research …” and, in fact, he argues, for many versions of the IAT, the test-retest reliability “is as high as (or higher than) that for self-administered blood-pressure readings” (p. 2, citing Brody, Veit, & Rau, 1999).

In terms of validity, Lane et al., (2007) reviewed the various assessments of the IAT and found solid evidence for construct validity, mixed results for convergent validity, and evidence of discriminant validity.[12] (See also Nosek et al., 2007.) The discussion of predictive validity has produced the battle of the meta-analyses with one team claiming that the IAT *does* predict

[10]Importantly, some of this discourse is not just about the IAT, per se, but rather about the challenges associated with measuring implicit cognitions more generally (see DeSchryver, 2018).

[11] Commenting on the findings of some studies that show low test-retest reliability, Payne et al. (2017) speculate that it may not be that the IAT measure is at fault; it may be that implicit bias is not as stable as some have assumed. Instead, implicit biases may be impacted significantly by context.

[12]They found that discriminant validity findings varied by the attitude to which the implicit association was compared.

behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Greenwald et al., 2015; Jost, Rudman, Blain, Carney, Dasgupta, Glaser, & Hadin, 2009; Kurdi, Seitchik, Axt, Carroll, Karapetyan, Kaushik,…& Banaji, 2019) another saying that it *does not* (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). Kurdi et al., (2019, p. 1), suggest that the answer is somewhere in the middle of these two polar extremes. The issue is not *whether* the IAT predicts behavior, but rather identifying the "conceptual and methodological conditions" under which the IAT predicts it.

Some researchers have commented, in particular, on the reliability and validity of IAT scores that are aggregated, as is the case in the current study. Jost (2019) observed that the IAT is "highly reliable at the aggregate level" and Payne et al., (2017) claim that aggregate scores may be more valid than individual ones since they are better understood to represent situational concept accessibility rather than stable individual traits. That said, continued research should focus on whether and to what extent individual-level bias on the part of criminal justice actors impacts individual-level decisions. Knowing the extent to which bias emanates from the individual, from the aggregate/culture, or from an interaction of the two would have important implications for interventions to reduce biased behavior and thus reduce disparities in the criminal justice system.

A problem with the measures of both explicit bias and implicit bias is the fact that individuals self-select to participate in the assessments. The individuals participating in these assessments are not necessarily representative of the population as a whole and individuals can and do participate in the assessments more than once. As noted above, this issue is countered by the facts that: (1) Project Implicit (PI) respondents show patterns of bias that are similar to that of nationally representative samples, (2) research has demonstrated that geographically-aggregated implicit bias measures have predicted expected outcomes, and (3) although the scores for the counties might not reflect the absolute level of racial bias within them, what is important for this study is the relative levels of explicit and implicit bias for the

counties, compared to the others. Arguably, confidence can be placed in the findings regarding the relationship between county-level bias and jail population disparities control if the selection bias operates similarly across counties. Notwithstanding the defense above to the use of PI data, as set forth in the description of the methods, we went further to reduce the potential risk of a non-random sample by using county-level measures estimated via Multilevel Regression with Poststratification (MRP).

Our results are correlational. Like most cross-sectional research studies, omitted variable bias is a potential concern. Although we controlled for a number of socioeconomic covariates, many measures of community context may be relevant to both bias and incarceration. For instance, the degree of urbanization and segregation/integration likely contributes to bias, crime, and punishment, and may contribute to perceptions of symbolic racial threat (à la group threat theories) or decreased racial prejudice (à la group contact theories). Other than implicit and explicit racial associations, attitudes were not measured. Given that conservatism, social dominance orientation, and right-wing authoritarianism are associated with both racial bias and punitiveness, it is possible that these unmeasured regional attitudes are responsible for the observed relationships, rather than bias *per se*.

Moreover, racial disparities in jail populations could be the *cause*, not the *consequence,* of individual or cultural racial attitudes. Indeed, it is likely that institutionalized, racialized patterns of criminal justice are observed and internalized, contributing to conscious or subconscious racial stereotypes and racial bias. Further, it is also probable that the relationship between bias and disparate representation in the criminal justice system is reciprocal. There could exist a positive feedback loop wherein higher levels of Black incarceration lead to higher levels of bias, and higher levels of bias lead to higher levels of Black incarceration. Future research should explore these potential non-recursive relationships.

Conclusion

Community-level implicit and explicit biases observed in this study demonstrate significant, substantive, and meaningful relationships with the racial composition of local jails. If these biases are both a cause and consequence of economic, historical, and cultural social structures and social systems, and if the general public is understood to contribute to criminal justice outcomes by invoking that system and democratically influencing punishment policy, then it is little wonder that training and policy tweaks within criminal justice agencies have demonstrated rather limited success. Reducing racial bias and its concomitant incarceration disparities will require radical changes to economic systems, educational systems, neighborhoods, and cultural products, both nationally and regionally. To the extent that these systemic changes remain out of reach, racial bias and racially disparate incarceration are likely to persist.

Declaration of Competing Interests:

None

Acknowledgments:

The authors thank John Cochran for his review of a manuscript draft and helpful input to strengthen it.

References

American Bar Association. (n.d.). Implicit Bias Initiative. Retrieved June 3, 2019, from https://www.americanbar.org/groups/litigation/initiatives/task-force-implicit-bias/

Amodio, D M., & Mendoza, S.A. (2010). Implicit intergroup bias: Cognitive, affective, and motivational underpinnings. In B. Gawronski (Ed.), *Handbook of implicit social cognition: Measurement, theory, and applications* (353–374). New York: The Guilford Press.

Angrist, J.D., & Pischke, J.S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*(3), pp. 668–688. https://doi.org/10.3758/s13428-013-0410-6

Blumstein, A., & Cohen, J. (1973). Theory of the stability of punishment. *Journal of Criminal Law and Criminology*, 64(2), 198–207.

Brody, S., Veit, R., & Rau, H. (1999). Four-year test-retest reliability of self-measured blood pressure. *Archives of Internal Medicine*, 169, 1007–1008.

Cameron, A. C., & Trivedi, P.K. (2013). *Regression Analysis of Count Data*. Cambridge University Press.

Casey, PM. Warren, R. K., Cheesman, F.L., & Elek, J.K. (2013). Addressing implicit bias in the courts. *49 Court Review* 64. Available at https://heinonline.org/HOL/LandingPage?handle=hein.journals/ctrev49&div=12&id=&page=).

Colligan, C. (2022). County passes motion to decriminalize bike violations. *Daily Bruin*, Feb 8. Found at: https://dailybruin.com/2022/02/08/la-county-passes-motion-to-decriminalize-bike-violations

Concannon, C. & Na, C. (2024). Examining racial and ethnic disparity in prosecutor's bail requests and downstream decision-making. *Race and Social Problems*, 16, 1–18. https://doi.org/10.1007/s12552-022-09385-0

Correll, J., Hudson, S.M., Guillermo, S., & Ma, D.S. (2014). The police officer's dilemma: A decade of research on racial bias in the decision to shoot. *Social and Personality Psychology Compass*, *8*(5), 201–213.

Cunningham, W.A., Preacher, K J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*(2), 163–170. https://doi.org/10.1111/1467-9280.00328

DeSchryver, M. (2018). *A psychometric analysis of choice reaction time measures*.

(Unpublished doctoral dissertation) Ghent University, Netherlands.

Dovidio, J.F. & Gaertner, S.L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychology Science, 11(4), 315–319.*

Downes, M., & Carlin, J.B. (2020). Multilevel regression and poststratification versus survey sample weighting for estimating population quantities in large population health studies. *American Journal of Epidemiology*, *189*(7), 717–725. https://doi.org/10.1093/aje/kwaa053

Durose, M. R., & Langton, L. (2013). *Police behavior during traffic and street stops, 2011* (Bureau of Justice Statistics Publication No. 242937). Washington, DC: U.S. Government Printing Office. Available at https://bjs.ojp.gov/library/publications/police-behavior-during-traffic-and-street-stops-2011

Esmer, Y., & Pettersson, T. (2007). *Measuring and mapping cultures: 25 years of comparative value surveys. Series: International Studies in Sociology and Social Anthropology, Volume 104.* Brill.

Ekstrom, PD., Lai, C.K., & Forestier, J.L. (2019). *Does regional racial bias predict disparities in police traffic stops?* [dataset]. OSF. https://osf.io/6r4zy/

Ekstrom, P.D., Le Forestier, J.M., & Lai, C.K. (2022). Racial demographics explain the link between racial disparities in traffic stops and county-level racial attitudes. *Psychological Science*, *33*(4), 497–509. https://doi.org/10.1177/09567976211053573

Federal Bureau of Investigation (2017). Crime in the United States, 2017. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s./2017/crime -in-the-u.s.-2017.

Fridell, L.A. (2017). *Producing bias-free policing: A science-based approach*. Springer.

Fridell, L., & Brown, S. (2015). Fair and impartial policing: A science-based approach. *The Police Chief, June,* 20–25.

Gawronski, B., Morrison, M., Phills, C.E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures. *Personality and Social Psychology Bulletin, 43*(3), 300–312. https://doi.org/10.1177/0146167216684131

Gayla, M. (2017, June 21). A federal court asks jurors to confront their hidden biases. Retrieved June 3, 2019, from The Marshall Project website: https://www.themarshallproject.org/2017/06/21/a-federal-court-asks-jurors-to-confront-their-hidden-biases

Gelman, A., Little, T.C., & Witter, M.S.D. (1997). *Poststratification into many categories using hierarchical logistic regression*. Available at http://www.stat.columbia.edu/~gelman/research/published/poststrat3.pdf

Gilliam, W.S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?* Report of the Yale Child Study Center. Available at: https://files-profile.medicine.yale.edu/documents/75afe6d2-e556-4794-bf8c-3cf105113b7c

Graham, S., & Lowery, B.S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior*, *28*(5), 483–504.

Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem and stereotypes. *Psychological Review*, 102(1), 4–27.

Greenwald, A.G., Banaji, M.R., & Nosek, B.A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, 108(4), 553–561.

Greenwald, A.G., Poehlman, T., Uhlmann, E.L., & Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41. https://doi.org/10.1037/a0015575

Griffin, J. (2024). Court orders Philly police to change quality-of-life enforcement to curb racial disparity. *The Philadelphia Inquirer*, March 21. Found at: https://www.inquirer.com/news/philadelphia-police-stop-and-frisk-racial-bias-20210604.html

Hardin, C.D., & Banaji, M.R., (2013). The nature of implicit prejudice: Implications for personal and public policy. *The Behavioral Foundations of Public Policy*, 13–31.

Hehman, E., Flake, J.K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*(4), 393–401.

Hetey, R.C., Hamedani, M.G., Markus, H.R., & Eberhardt, J. (2024). When the cruiser lights come on: Using the science of bias & culture to combat racial disparities in policing. Daedalus, Winter, 123–150.

Hester, R., & Hartman, T.K. (2017). Conditional race disparities in criminal sentencing: A test of the Liberation Hypothesis from a non-guidelines state. *Journal of Quantitative Criminology*, *33*(1), 77–100.

Hilbe, J.M. (2011). *Negative binomial regression*. Cambridge University Press.

Hilbe, J.M. (2014). *Modeling count data*. Cambridge University Press.

Hipp, J.R. (2011). Spreading the wealth: The effect of the distribution of income and race/ethnicity across households and neighborhoods on city crime trajectories. *Criminology*, *49*(3), 631–665.

Hoffman, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31(10), 1369–1385.

Huebner, B. M., & Bynum, T.S. (2008). The role of race and ethnicity in parole decisions. *Criminology*, 46(4), 907 – 938.

Jost, J.T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science*, *28*(1), 10–19. https://doi.org/10.1177/0963721418797309

Jost, J. T., Rudman, L.A., Blair, I.V., Carney, D.R., Dasgupta, N., Glaser, J., & Hardin, C.D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological A objections and executive summary of ten studies that no manager should ignore. *94 California Law Review*, 1063.

Kang, J. (2024). Little things matter a lot: The significance of implicit bias, practically and legally. *Daedalus*, Winter, 193–212. https://doi.org/10.1162/daed_a_02055

Kang, J., & Banaji, M. (2006). Fair measures: A behavioral realist revision of affirmative action. *California Law Review*, 94(4), 1063–1118.

Kang, J., Bennett, M., Carbado, D., & Casey, P. (2011). Implicit bias in the courtroom. *UCLA Law Review*, *59*(5), 1124–1186.

Kaplan, J. (2020). *Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data: Arrests by Age, Sex, and Race, 1974-2018*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-06-29. https://doi.org/10.3886/E102263V10

Kurdi, B., Seitchik, A.E., Axt, J.R., Carroll, T.J., Karapetyan, A., Kaushik, N., … Banaji, M.R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *The American Psychologist, 74(5)*, 569–586.https://doi.org/10.1037/amp0000364

Kutateladze, B.L. (2018). Tracing charge trajectories: A study of the influence of race in charge changes at case screening, arraignment, and disposition. *Criminology*, *56*(1), 123–153.

Lane, K.A., Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2007). Understanding and using the Implicit Association Test: Procedures and validity. In Wittenbrink, B. & Schwarz, N. (Eds). *Implicit Measures of Attitudes: Procedures and Controversies*, 59–102.

Lautenschlager, R. & Omori, M. (2019). Racial threat, social organization, and police use of force in communities of color. *Justice Quarterly*, 36, 1050–1071.

Lawler, C. (2023). Duluth police, NAACP together write new policy on traffic stops. Star Tribune, Nov 18. Found at: https://www.wdio.com/front-page/top-stories/new-changes-to-duluth-police-department-traffic-stops-in-2024/

Leitner, J.B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016a). Blacks' death rate due to circulatory diseases is positively related to Whites' explicit racial bias. *Psychological Science, 27*(10), 1299–1311. https://doi.org/10.1177/0956797616658450

Leitner, J.B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016b). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine*,170, 220–227. https://doi.org/10.1016/j.socscimed.2016.10.007

Leitner, J.B., Hehman, E., & Snowden, L.R. (2018). States higher in racial bias spend less on disabled Medicaid enrollees. *Social Science & Medicine*, *208*, 150–157.

Levinson, J.D., Cai, H., & Young, D. (2010). Guilty by implicit racial bias: The guilty/not guilty Implicit Association Test. *Ohio State. Journal of Criminal Law*, *8*, 187.

Loeber, R., Farrington, D.P., Hipwell, A.E., Stepp, S.D., Pardini, D., & Ahonen, L. (2015). Constancy and change in the prevalence and frequency of offending when based on longitudinal self-reports or official records: Comparisons by gender, race, and crime type. *Journal of Developmental and Life-Course Criminology*, *1*(2), 150–168.

Mears, D.P., Cochran, J.C., & Lindsey, A.M. (2016). Offending and racial and ethnic disparities in criminal justice. *Journal of Contemporary Criminal Justice, 32*(1), 78–103. https://doi.org/10.1177/1043986215607252

Miller, C.T., Varni, S.E., Solomon, S.E., DeSarno, M.J., & Bunn, J. Y. (2016). Macro-level implicit HIV prejudice and the health of community residents with HIV. *Health Psychology*, *35*(8), 807–815.

Minton, T.D. (2021). Correctional Populations in the United States, 2019 – Statistical Tables. *Statistical Tables*. Available at https://bjs.ojp.gov/library/publications/correctional-populations-united-states-2019-statistical-tables

Mitchell, G., & Tetlock, P. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In Lilienfeld, S.O. & Waldman, I.D. (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions,* pp. 164–195. Wiley

Blackwell. https://doi.org/10.1002/9781119095910.ch10

Nagai, A. (n.d.). The Implicit Association Test: Flawed science tricks americans into believing they are unconscious racists. Retrieved June 3, 2019, from The Heritage Foundation website: https://www.heritage.org/science-policy/report/the-implicit-association-test-flawed-science-tricks-americans-believing-they

Nicklas, M. (2012). Austin police adopt stricter rules for searches, but enforcement is key. *Reporting Texas*, December 17. Found at: https://www.reportingtexas.com/austin-police-consent-searches-written-vehicle/

Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. Bargh (ed.) *Frontiers of social psychology. Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York: Psychology Press.

Nosek, B.A., Smyth, F.L., Sriram, N., Lindner, N.M., Devos, T., Ayala, A., … Greenwald, A.G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10593–10597.

Orchard, J., & Price, J. (2017). County-level racial prejudice and the Black-White gap in infant health outcomes. *Social Science & Medicine*, *181*, 191–198.

Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, *108*(4), 562–571.

Oude Maatman, F. (2017). *Implicit assumptions: A case study on the IAT controversy*. (Unpublished doctoral dissertation). Radboud University Nijmegen, Nijmegen, Netherlands.

Payne, B.K., & Vuletich, H.A. (2018). Policy insights from advances in implicit bias research. *Behavioral and Brain Sciences*, 5(1), 49 – 56.

Payne, B.K., Vuletich, H.A., & Lundberg, K.B (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248.

Pena, L. (2024). New SF policy aims to limit ‘pretextual’ traffic stops, tackle racial bias. *ABC News, Bay Area.* Found at: https://abc7news.com/sf-police-traffic-stops-crime-san-francisco-commission/14462590/

Pettigrew, T.F., & Tropp, L R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology,* 90(5), 751–783.

Petty, R.E., Fazio, R.H., & Brinol, P. (2012). *Attitudes: Insights from the New Implicit Measures*.

New York: Psychology Press.

Pinkston, K. (2015). The Black-White malleability gap in implicit racial evaluations: A nationally representative study. *The Journal of Social Psychology*, *155*(3), 189–203.

Rachlinski, J.J., Johnson, S.L., Wistrich, A.J., & Guthrie, C. (2009). Does unconscious bias affect trial judges? *The Notre Dame Law Review*, *84*(3), 1195–1246.

Rae, J.R., Skinner-Dorkenoo, A.L. … and Hewstone, M. (2022). *Social Psychological and Personality Science*, 13(2), 476–489. https://doi.org/10.1177/19485506211027756

Ratliff, K.A. & Smith, C.T. (2024). The implicit association test. *Daedalus*, Winter: 51 – 54. https://doi.org/10.1162/daed_a_0248.

Rezaei, A. R. (2011). Validity and reliability of the IAT: Measuring gender and ethnic stereotypes. *Computers in Human Behavior*, 27, 1937–1941. https://doi.org/10.1016/j.chb.2011.04.018

Richardson, L.S., & Goff, P.A. (2012). Implicit racial bias in public defender triage. *Yale Law Journal*, *122*(8), 2626–2649.

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias*. Proceedings of the National Academy of Sciences, 116*(17), 8255–8260. https://doi.org/10.1073/pnas.1808307116

Sah, S., Robertson, C., & Baughman, S.B. (2015). Binding prosecutors to defendants' race: A policy proposal to reduce unconscious bias in the criminal justice system. *Behavioral Science & Policy*, 1(2),69–76.

Salman, J., Le Coz, E., & Johnson, E. (2016). Florida's broken sentencing system: Designed for fairness, it fails to account for prejudice. *Sarasota Herald-Tribune*. Retrieved from http://projects.heraldtribune.com/bias/sentencing.

Sampson, R. J., & Lauritsen, J.L. (1997). Racial and ethnic disparities in crime and criminal justice in the United States. *Crime and Justice*, 21, 311 – 374.

Sampson, R.J, Morenoff, J.D, & Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health, 95*(2), 224–232. https://doi.org/10.2105/ajph.2004.037705

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*(5328), 918–924.

Sernoffsky, E. (2019). SF DA Gascón launching tool to remove race when deciding to charge suspects. *San Francisco Chronicle*, June 12. Found at: https://www.sfchronicle.com/crime/article/SF-DA-Gasc-n-launching-tool-to-remove-race-

when-13971721.php

Sim, J.J., Correll, J., & Sadler, M.S. (2013). Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in a decision to shoot. *Personality and Social Psychology Bulletin*, 39(3), 291–304.

Skeem, J L., Montoya, L., & Lowenkamp, C. (2023). Place matters: Racial disparities in pretrial detention recommendations across the U.S. *Federal Probation, 86(3), 5–10.*

Skinner-Dorkenoo, A.L., George M, Wages III, J.E., Sánchez, S., & Perry, S. P. (2023). A systemic approach to the psychology of racial bias within individuals and society. *Nature Reviews Psychology*, *2*(7), 392–406.

Smith, R.J., & Levinson, J.D. (2012). *The impact of implicit racial bias on the exercise of prosecutorial discretion*. Seattle University Law Review, 35(3), 795–826.

Staats, C. (2013). *State of the science: Implicit bias review.* Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity. Available https://search.issuelab.org/resource/state-of-the-science-implicit-bias-review-2013.html

Staats, C., Capatosto, K., Wright, R.A., & Jackson, V.W. (2016). *State of the science: Implicit bias review, 2016 Edition*. Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity. Available at https://search.issuelab.org/resource/state-of-the-science-implicit-bias-review-2016.html

Tonry, M. & Melewski, M. (2008). The malign effects of drug and crime control policies on Black Americans. *Crime and Justice*, 37, 1 – 47.

U.S. Census Bureau. (n.d.). Tables B02001, B03003, B19083, S1201, S1903, and S2301 [Data set]. *American Community Survey 3-Year Estimates*. Retrieved from https://www.census.gov/programs-surveys/acs

U.S. Sentencing Commission. (2018). Demographic differences in sentencing. *Federal Sentencing Reporter, 30*(3), 212–229. https://doi.org/10.1525/fsr.2018.30.3.212

Valencia, N. (2023). Memphis leaders pass measure that ends police stops for minor infractions, three months after Tyre Nichols' death. CNN News, April 12. Found at: https://www.cnn.com/2023/04/12/us/memphis-pretext-stops-tyre-nichols/index.html https://www.cnn.com/2023/04/12/us/memphis-pretext-stops-tyre-nichols/index.html

Vera Institute of Justice. (2018). Incarceration Trends Dataset: County-level jail data (1970-2015) and prison data (1983-2015). (Version 1.0) [Data set]. GitHub. https://github.com/vera-institute/incarceration-trends/

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls*. International Journal of Forecasting*, *31*(3), 980–991. https://doi.org/10.1016/j.ijforecast.2014.06.001

White, M.D., & Fradella, H.F., (2016). *Stop and Frisk: The Use and Abuse of a Controversial Policing Tactic*. New York: NYU Press.

Wilson, A. (2023). Probation and parole revocations: Reviewing over two decades of revocation decisions in the United States. *Criminal Justice Review*, https://doi.org/10.1177/07340168231170335.

Wistrich, A., & Rachlinski, J. (2017) Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science*, 13, 203–229.

Xu, F.K., Nosek, B.A., Greenwald, A.G., Ratliff, K., Bar-Anan, Y., Umansky, E., … Axt, J. (2019). *Project Implicit Demo Website Datasets*. https://doi.org/10.17605/OSF.IO/Y9HIQ

Zeng, X., Xu, L., & Xiao, X. (2023). Negative implicit in-group stereotypes of Chinese male drug abusers: Evidence from ERP. *Current Psychology*, 42, 23861–23873.

Zeng, Z., & Minton, T.D. (2021). *Jail inmates in 2019* (NCJ 255608). Bureau of Justice Statistics. Available at https://bjs.ojp.gov/library/publications/jail-inmates-2019

Zerhouni, O., Rougier, M., & Muller, D. (2016). “Who (really) is Charlie?” French cities with lower implicit prejudice toward Arabs Demonstrated larger participation rates in Charlie Hebdo rallies [“Qui est (Vraiment) Charlie?” Les Villes Françaises à plus Faible niveau de Préjugés Implicites envers les Maghrébins ont davantage Participé aux rassemblements de Charlie Hebdo]. *International Review of Social Psychology*, *29*(1). Retrieved from https://www.rips-irsp.com/article/10.5334/irsp.50/

Zhang, X., Holt, J.B., Lu, H. Wheaton, A.G., Ford, E.S., Greenlund, K.J., & Croft, J.B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American Journal of Epidemiology*, *179*(8), 1025–1033. https://doi.org/10.1093/aje/kwu018

Fig. 1 *White Incarceration Rate, County Jails, 2011-2015*

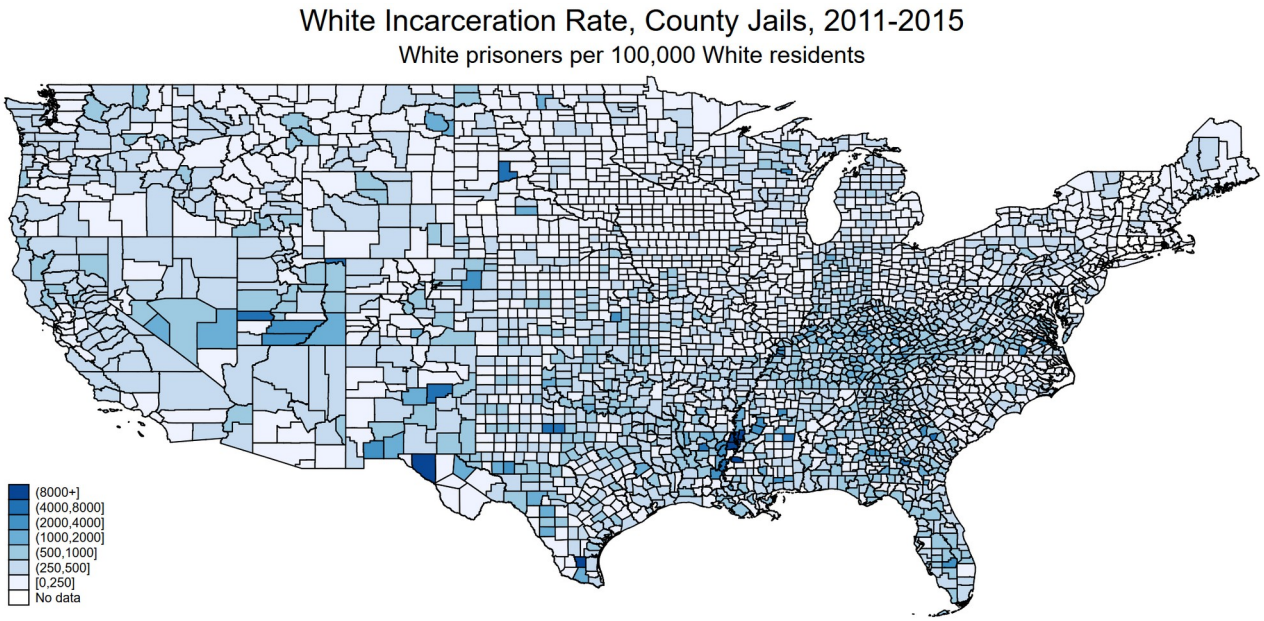

Fig. 2 *Black Incarceration Rate, County Jails, 2011-2015*

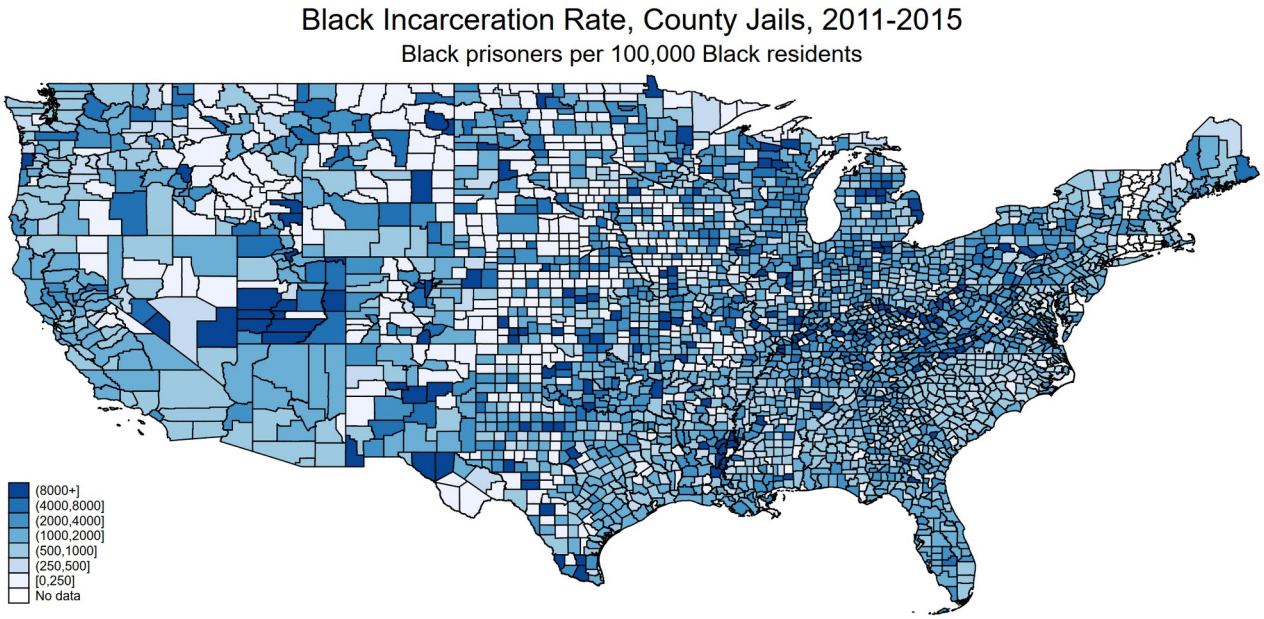

Table 1: Descriptive Statistics of Study Variables, N = 2,826

Variable	Mean	S.D.	Min	Max
Black Prisoners*	443.73	1,563.98	0	30,950
White Prisoners*	576.18	1,057.40	0	17,463
Black Population 15+*	47,663.26	223,903	10	6,750,020
White Population 15+*	227,749.10	535,874	986	9,729,438
Implicit Bias (MRP)	.338	.064	-.024	.417
Explicit Bias (MRP)	.443	.341	-1.542	1.078
Black Employment Ratio	43.64	20.67	0	100
Black Marriage Rate	28.08	16.22	0	100
Black Poverty Rate	32.13	21.04	0	100
Black Arrest Rate (p/1,000)	129.24	162.39	0	5,407
White Employment Ratio	54.66	7.80	21.9	77.1
White Marriage Rate	55.08	5.30	17.2	74.3
White Poverty Rate	13.36	5.07	0.7	40.3
White Arrest Rate (p/1,000)	57.47	148.08	0	4,581
Index Crimes*	13,502.54	48,210.93	0	1,088,822

**Notes: 5-year totals, 2011-2015*

Table 2: Zero-inflated Negative Binomial Regression of Black and White Prisoners in County Jails

	(1) Black Prisoners	(2) White Prisoners	(3) Black Prisoners	(4) White Prisoners
Implicit Bias	1.261***	.792***		
	(.060)	(.052)		
Explicit Bias			1.241***	.772***
			(.053)	(.049)
Black Employment Ratio	.989*	.999	.988*	.999
	(.005)	(.001)	(.005)	(.001)
Black Marriage Rate	.993	.999	.993	.999
	(.004)	(.001)	(.004)	(.001)
Black Poverty Rate	1.004	1.000	1.003	1.001
	(.003)	(.002)	(.003)	(.002)
Black Arrest Rate	1.000	1.000	1.000	1.000
	(.000)	(.000)	(.000)	(.000)
White Employment Ratio	1.015	.974***	1.016	.977***
	(.013)	(.006)	(.013)	(.006)
White Marriage Rate	1.069***	1.022**	1.068***	1.025**
	(.018)	(.008)	(.019)	(.008)
White Poverty Rate	1.054*	1.028***	1.055*	1.032***
	(.022)	(.009)	(.023)	(.008)
White Arrest Rate	1.000	1.002**	1.000	1.002**
	(.000)	(.001)	(.000)	(.001)
Exposure[a]	1	1	1	1
Negative Binomial Dispersion Parameter	1.237	.534***	1.247	.527***
	(.177)	(.057)	(.181)	(.057)
Inflate				
Index Crimes	0.999*	1.000*	0.999*	1.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Total Population	1.000**	1.000	1.000*	1.000
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,826	2,826	2,826	2,826
AIC	31,096	37,208	31,122	37,168
BIC	31,180	37,291	31,205.5	37,251
Log-likelihood	-15,534	-18,590	-15,547	-18,570
Wald χ^2	114.9	105.7	131.8	110.1

*Notes: Incident Rate Ratios. Standard errors in parentheses. [a]Exposure term for Models 1 and 3 is Black population-years. Exposure term for Models 2 and 4 is White population-years. *p < .05, **p < .01, ***p < .001*

Table 3: Multilevel Negative Binomial Regression of Black and White Prisoners in County Jails

	(5)	(6)	(7)	(8)
	Black Prisoners	White Prisoners	Black Prisoners	White Prisoners
Implicit Bias	1.236***	.851***		
	(.036)	(.018)		
Explicit Bias			1.224***	.854***
			(.035)	(.017)
Black Employment Ratio	.995***	.999	.995***	.999
	(.001)	(.001)	(.001)	(.001)
Black Marriage Rate	.998	.998	.998	.998
	(.002)	(.001)	(.002)	(.001)
Black Poverty Rate	1.004**	1.002**	1.004**	1.002**
	(.001)	(.001)	(.001)	(.001)
Black Arrest Rate	1.001***	1.001***	1.001***	1.001***
	(.000)	(.000)	(.000)	(.000)
White Employment Ratio	1.009	.982***	1.009	.982***
	(.005)	(.003)	(.005)	(.003)
White Marriage Rate	1.046***	1.014***	1.046***	1.014***
	(.005)	(.004)	(.005)	(.004)
White Poverty Rate	1.020**	1.018***	1.020**	1.018***
	(.007)	(.005)	(.007)	(.005)
White Arrest Rate	1.000	1.001***	1.000	1.001***
	(.000)	(.000)	(.000)	(.000)
Exposure[a]	1	1	1	1
Negative Binomial Dispersion Parameter	1.330***	.619***	1.331***	.619***
	(.036)	(.017)	(.036)	(.017)
State Random Effects Variance (S.D.)	1.257***	1.154***	1.262***	1.152***
	(.069)	(.039)	(.070)	(.038)
N	2,826	2,826	2,826	2,826
AIC	30,969	37,300	30,974	37,301
BIC	31,040	37,372	31,045	37,372
Log-likelihood	-15,472	-18,638	-15,475	-18,638
Wald χ^2	243.7	245.2	239.7	244.8

*Notes: Incident Rate Ratios. Clustered standard errors in parentheses. [a]Exposure term for Models 5 and 7 is Black population-years. Exposure term for Models 6 and 8 is White population-years. *p < .05, **p < .01, ***p < .001*

Appendix: Supplemental Data Analysis

In the main manuscript, we report multilevel negative binomial models and zero-inflated negative binomial models. In this supplement, we report several additional model specifications: (a) standard Poisson regression models; (b) fixed-effects Poisson regression models; (c) random-effects Poisson regression models; and (d) multilevel negative binomial models that restrict the sample to only those counties that are at least 2% Black and had at least 50 IAT respondents. Fixed-effects negative binomial models are not explored, given that they are problematic and do not effectively control for county-level fixed effects (Wooldridge, 2010).

In every specification herein, increases in county-level anti-Black implicit bias and explicit bias were associated with a higher number of population-adjusted Black prisoners, consistent with the models reported in the main manuscript. However, several specifications diverge from the main models for White prisoners, where they generally show non-significant relationships between bias and White prisoner counts. Nonetheless, model fit is generally poorer in the supplemental models.

Table A: Poisson Regression of Black and White Prisoners in County Jails

	Black Prisoners	**White Prisoners**	**Black Prisoners**	**White Prisoners**
Implicit Bias	1.112***	.948		
	(.030)	(.026)		
Explicit Bias			1.119***	.951
			(.032)	(.026)
Black Employment Ratio	(.004)	(.001)	(.003)	(.001)
	1.005	1.001	1.005	1.001
Black Marriage Rate	(.005)	(.001)	(.006)	(.001)
	1.008*	1.000	1.008*	1.000
Black Poverty Rate	(.004)	(.001)	(.004)	(.001)
	1.002***	1.000	1.002**	1.000
Black Arrest Rate	(.001)	(.000)	(.001)	(.000)
	.991	.969***	.990	.969***
White Employment Ratio	(.006)	(.003)	(.006)	(.003)
	1.005	1.025***	1.002	1.026***
White Marriage Rate	(.007)	(.004)	(.007)	(.005)
	1.029**	1.039***	1.026**	1.040***
White Poverty Rate	(.009)	(.005)	(.010)	(.005)
	.998**	1.000	.998*	1.000
White Arrest Rate	(.001)	(.000)	(.001)	(.000)
	1.112***	.948		
Exposure Term[a]	1	1	1	1
Observations	2826	2826	2826	2826
AIC	541183.2	562726.8	540707.7	563110.4
BIC	541242.6	562786.3	540767.2	563169.9
ll	-270581.6	-281353.4	-270343.9	-281545.2
chi2	179.0	507.7	178.3	509.2

*Notes: Incident Rate Ratios; Clustered standard errors in parentheses. [a]Exposure term for Black Prisoner models is Black population-years. Exposure term for White Prisoner models is White population-years. *p < .05, **p < .01, ***p < .001*

Table B: State-Fixed-Effects Poisson Regression of Black and White Prisoners in County Jails

	Black Prisoners	**White Prisoners**	**Black Prisoners**	**White Prisoners**
Implicit Bias	1.150***	1.007		
	(.043)	(.046)		
Explicit Bias			1.149***	1.016
			(.039)	(.043)
Black Employment Ratio	.995	.997**	.995	.997**
	(.004)	(.001)	(.004)	(.001)
Black Marriage Rate	1.000	1.000	1.000	1.000
	(.004)	(.001)	(.004)	(.001)
Black Poverty Rate	1.014***	1.002*	1.013***	1.002*
	(.003)	(.001)	(.003)	(.001)
Black Arrest Rate	1.002*	1.000*	1.002	1.000*
	(.001)	(.000)	(.001)	(.000)
White Employment Ratio	.998	.974***	.999	.974***
	(.008)	(.004)	(.008)	(.004)
White Marriage Rate	1.014	1.011*	1.012	1.010
	(.008)	(.005)	(.008)	(.005)
White Poverty Rate	1.028**	1.017*	1.026**	1.016*
	(.009)	(.007)	(.009)	(.006)
White Arrest Rate	.998	1.000	.998	1.000
	(.001)	(.000)	(.001)	(.000)
Exposure Term[a]	1	1	1	1
Observations	2826	2826	2826	2826
AIC	413368.5	375859.9	412627.8	375648.5
BIC	413422.0	375913.4	412681.3	375702.0
ll	-206675.3	-187920.9	-206304.9	-187815.2
chi2	380.6	294.7	399.4	298.4

*Notes: Incident Rate Ratios; Clustered standard errors in parentheses. [a]Exposure term for Black Prisoner models is Black population-years. Exposure term for White Prisoner models is White population-years. *p < .05, **p < .01, ***p < .001*

Table C: Random-Effects Poisson Regression of Black and White Prisoners in County Jails

	Black Prisoners	**White Prisoners**	**Black Prisoners**	**White Prisoners**
Implicit Bias	1.150***	1.007		
	(.043)	(.046)		
Explicit Bias			1.149***	1.016
			(.040)	(.043)
Black Employment Ratio	.995	.997**	.995	.997**
	(.004)	(.001)	(.004)	(.001)
Black Marriage Rate	1.000	1.000	1.000	1.000
	(.004)	(.001)	(.004)	(.001)
Black Poverty Rate	1.014***	1.002*	1.013***	1.002*
	(.003)	(.001)	(.003)	(.001)
Black Arrest Rate	1.002*	1.000*	1.002	1.000*
	(.001)	(.000)	(.001)	(.000)
White Employment Ratio	.998	.974***	.999	.974***
	(.008)	(.004)	(.008)	(.004)
White Marriage Rate	1.014	1.011*	1.012	1.010
	(.008)	(.005)	(.008)	(.005)
White Poverty Rate	1.028**	1.017*	1.026**	1.016*
	(.009)	(.007)	(.009)	(.006)
White Arrest Rate	.998	1.000	.998	1.000
	(.001)	(.000)	(.001)	(.000)
Exposure Term[a]	1	1	1	1
Random-Effects Parameter	.240	.135	.248	.136
	(3.084)	(1.304)	(3.316)	(1.321)
Observations	2826	2826	2826	2826
AIC	414283.1	376809.2	413544.0	376597.9
BIC	414348.5	376874.6	413609.4	376663.3
ll	-207130.6	-188393.6	-206761.0	-188288.0
chi2	3495.5	17469.0	3504.8	18473.8

*Notes: Incident Rate Ratios; Clustered standard errors in parentheses. [a]Exposure term for Black Prisoner models is Black population-years. Exposure term for White Prisoner models is White population-years. *p < .05, **p < .01, ***p < .001*

Table D: Multilevel Negative Binomial Regression of Black and White Prisoners in County Jails, Within-county $N_{IAT} \geq 50$ and Black Population ≥2%

	Black	**White**	**Black**	**White**
Implicit Bias	1.175***	.949*		
	(.036)	(.024)		
Explicit Bias			1.163***	.960
			(.035)	(.024)
Black Employment Ratio	.999	.998	.999	.999
	(.003)	(.002)	(.003)	(.002)
Black Marriage Rate	.993	.995	.993	.994
	(.004)	(.003)	(.004)	(.003)
Black Poverty Rate	1.005	1.001	1.005	1.001
	(.003)	(.002)	(.003)	(.002)
Black Arrest Rate	1.003***	1.000	1.003***	1.000
	(.000)	(.000)	(.000)	(.000)
White Employment Ratio	.995	.969***	.996	.969***
	(.006)	(.005)	(.006)	(.005)
White Marriage Rate	1.017*	1.019**	1.016*	1.018**
	(.007)	(.006)	(.007)	(.006)
White Poverty Rate	1.010	1.013	1.010	1.012
	(.010)	(.008)	(.010)	(.008)
White Arrest Rate	.998	1.004***	.999	1.004***
	(.001)	(.001)	(.001)	(.001)
Exposure Term[a]	1	1	1	1
Dispersion Parameter	.428***	.280***	.429***	.280***
	(.021)	(.014)	(.022)	(.014)
State Random Effects Variance	1.079**	1.110***	1.081**	1.112***
	(.027)	(.031)	(.028)	(.031)
Observations	815	815	815	815
AIC	11735.4	12307.9	11738.1	12309.4
BIC	11791.8	12364.3	11794.5	12365.9
ll	-5855.7	-6141.9	-5857.1	-6142.7
chi2	153.2	205.2	150.2	203.3

*Notes: Incident Rate Ratios; Clustered standard errors in parentheses. [a]Exposure term for Black Prisoner models is Black population-years. Exposure term for White Prisoner models is White population-years. *p < .05, **p < .01, ***p < .001*