Measuring Racial Disparity in Local and County Police Arrests

Beth Redbird
Assistant Professor of Sociology and IPR Fellow
Northwestern University

Kat Albrecht
Department of Sociology and Pritzker School of Law
Northwestern University

Version: June 18, 2020

DRAFT
Please do not quote or distribute without permission.

ABSTRACT

Racial disparity in arresting behavior is a significant concern for people of color. It also delegitimizes law enforcement, increases tension between police and citizens, and can even increase crime. Efforts at police reform stall, in part because racial disparity in policing was previously unmeasurable. The authors present three new measures of racial disparity in arrest, measured across more than 13,000 agencies nationwide, allowing for reliable analysis of disparity across time and geographic space. These measures are validated against Department of Justice law enforcement misconduct investigations, the "gold standard" for assessing racial disparity in policing.

INTRODUCTION

Racial bias in police arrests is not only problematic for people of color, it is also a threat to the legitimacy of law enforcement agencies and a danger to fundamental democratic values (Goff and Kahn 2012, Tyler and Huo 2002). The public expects the police to be fair, impartial, effective, and restrained in the pursuit of justice (Skogan and Frydl 2004). Increases in fairness and impartiality increase the likelihood that police action will be accepted and supported by the public, and may itself decrease crime rates (Skogan and Frydl 2004).

Policing is often the most visible form of state interaction (Goff and Kahn 2012, Skogan and Frydl 2004). Police also have a significant amount of discretion on the decision to ignore, warn, sanction, or ultimately arrest (Skogan and Frydl 2004). This discretion can create disparate outcomes for different racial groups. Because arrest is the first stage of the criminal justice pipeline, bias created by police action continues throughout the whole criminal justice process (de Lint 2003, Hartman and Belknap 2003).

In 1994, in response to the Rodney King controversy, Congress passed Violent Crime Control and Law Enforcement Act[1] which gave the Department of Justice (DOJ) Civil Rights Division new authority to investigate and litigate law enforcement agencies with "a pattern or practice of [problematic] conduct" (Rushin 2014, U.S. Department of Justice 2017).

This was the authority used by DOJ officials to investigate police conduct in Ferguson, Missouri. After 100 person-days of onsite investigating, the DOJ concluded:

> *"The harms of Ferguson police and court practices are borne disproportionately by African Americans, and there is evidence that this is due in part to intentional discrimination on the basis of race."* (Justice 2015:4)

The massive data collection effort by DOJ investigators highlights one of the significant problems facing researchers of race and crime: there is currently no well-developed measure of racial bias in arrests. Without valid measurement of racial bias, policy initiatives designed to reform police practices easily stall (Skogan 2008, Walker 2012). Policy-makers have difficulty mandating change without an understanding of the policies or procedures that reduce or exacerbate police bias (Walker 2007).

This paper discusses the development of three measures of racial disparity at the *agency* level for Black, Asian, and American Indian / Alaskan Native (AIAN) arrests. The first measure developed is a simple risk ratio and provides a comparative baseline to assess racial disparities in arrest. However, geographic and demographic factors affect the rate of crime for an area's racial populations in different ways, making comparisons between racial groups difficult. As such, the second measure controls for group differences in factors related to the underlying commission of crime. Finally, the legal definitions of certain crimes differ across jurisdictions, limiting the

---

[1] Originally passed as 42 U.S.C. § 14141, recoded as 34 U.S.C. § 12601.

comparability of crime rates across geographies or over time. Thus, our final measure adjusts for geographic and temporal differences in the legal definition of crime. Together, these measures provide a suite of options for understanding the causes and consequences of nationwide changes in policing and law enforcement behavior.

MEASURING ARRESTS AND ARRESTABLE POPULATIONS

DATA ON ARRESTS

In general, crime rates are measured in one of three ways. First, some studies utilize the number of incidents reported to the police (for example see Osgood 2000). However, not all crimes are reported, and even fewer ultimately conclude in arrest. So, while number of incidents reported can provide a conservative estimate of crime rate, it provides no useful measure of racial difference in crime commission or institutional bias at entrance into the criminal justice system.

Second, other studies make use of victim data, such as the National Crime Victimization Survey (NCVS) administered by the Bureau of Justices Statistics (for examples, see Bachman 1998, Hashima and Finkelhor 1999, O'Brien 1996, Teplin et al. 2005). While such studies do not rely solely on the willingness of victims to make police reports, they still face problems with memory loss, willingness to be honest with researchers, ambiguities in questions, and issue salience (Skogan 1982). Additionally, not all crimes have a reportable victim. For instance, drug use, vagrancy, runaways, curfew violations, and illegal gambling, which in 2015 made up 22.6 percent of arrests, are likely missed by measures of victimization. Finally, this data cannot be used to measure racial differences in criminal justice contacts, as many victims do not know the race of the offender or whether the offender has eventual police contact.

The third method of crime rate measurement, used here, is the number of arrests made by police agencies. These data come from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) Program, the *primary* source of national data on criminal arrests, which compiles monthly arrest data from participating state and local law enforcement agencies. The data are publicly available through the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan. The UCR is a voluntary reporting program in which law enforcement agencies submit arrest statistics to the FBI, or in some instances to a state agency which then aggregates and forwards to the FBI (Maltz and Targonski 2002).

UCR data have been soundly and appropriately criticized for numerous methodological problems, particularly as they relate to creating trend estimates for total number of crimes committed at the national, state, or county level.

Despite wide use, the FBI cautions this data generally cannot be used to measure crime rates. because UCR data:

"provide[s] no insight into the many variables that mold the crime in a particular town, city, county, state, region, or other jurisdiction. Consequently, these rankings lead to simplistic and/or incomplete analyses that often create misleading perceptions adversely affecting cities and counties, along with their residents" (Justice and Investigation 2010).

These data do not include crimes which are unreported or remain unsolved. As such, this measurement does not accurately portray a crime rate, though it provides reasonable estimates for the number of arrests made by law enforcement agencies (Maltz 1999), and can therefore provide

a starting point for measuring racial disparities in arrest. Additionally, crime reported to police may be influenced by factors that also effect the underlying causes of crime, such as citizen trust, social inequality, and politicking practices (Maltz 1999). However, arrest data is also the only crime rate data source currently able to provide agency-level data on the race of the arrestee.[2]

The problem with UCR arrest data leads some researchers to conclude that measurement of racial bias in arrests may not be possible without some form of external regulation mandating reporting of racial behavior from agencies (Goff and Kahn 2012, Harmon 2013). Yet, since 2000 congress has regularly failed to pass legislation requiring municipal law enforcement agencies to collect demographic data in traffic stops (Ross and Parke 2009).[3]

In answer to this problem, this paper develops three measures of racial bias across nine demographic groups. Specifically, we developed a measure for Black, Asian, and American Indian arrests, and for adults, juveniles, and total arrests. Measures were developed for county, local, and tribal law enforcement agencies from 1999 to 2015.

The paper begins with a discussion of arrest data. It then proceeds to detail the construction of each measure, along with a discussion of the benefits and limitations of each measure. It concludes with an external validation test of each measure using agency investigation documents from the Department of Justice.

Uniform Crime Reporting Data

Data on arrests comes from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) Program from 1999 to 2015. UCR compiles monthly arrest data for participating state and local law enforcement agencies.[4] Each law enforcement agency participating in the UCR Program counts one arrest for each separate instance in which a person is arrested, cited, or summoned for an offense. The UCR Program divides offenses into two groups: Part I (which are primarily serious and violent crimes) and Part II crimes (everything else). Each month, contributing agencies submit information on: the number of Part I offenses known to law enforcement; offenses cleared by arrest or exceptional means; and age, sex, and race of persons arrested for each of the offenses. For Part II offenses, contributors provide only arrest data.

Some arrests involve multiple crimes arising from a single police enforcement action, termed a "multiple-offense situation". When reporting on situations involving more than one Part I offense, the arrest is classified by the reporting agency as the highest offense on the Part I schedule

---

[2] Because of the substantial data problems associated with UCR data, most studies of police discrimination rely on surveys that ask whether community members feel that their local police are fair and unbiased (Leiber, Nalla and Farnworth 1998, Stewart et al. 2009, Weitzer and Tuch 2004, Weitzer and Tuch 2005).

[3] E.g. Traffic Stops Statistics Act, 1997 and 2000; End Racial Profiling Act, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2013, 2015, 2016, 2017.

[4] U.S. territories and the District of Columbia are excluded.

hierarchy, and not as any other offense. For instance, an arrest that involves both burglary and assault would be classified as an assault for the purposes of reporting. The same rule applies for reporting of a multiple-offense arrest where the highest offense is Part I and the lowest offense is Part II. In contrast, a multiple-offense incident involving two Part II offenses is reported as *both* offenses (and thus recorded as multiple arrests). See the Uniform Crime Reporting Handbook (U.S. Department of Justice and Federal Bureau of Investigation 2004) for a list and definition of Part I and Part II offenses.

Reported arrest numbers are reviewed and processed by UCR Program Staff, and outlying reports are verified or corrected.[5] Agencies are also able to submit corrections to the previous month's arrest count in the following month.

Race of Arrestee

Along with a count of total arrests, agencies also report the race, gender, and age of arrested individuals. Racial categories were adopted from the Office of Federal Statistical Policy and Standards, U.S. Department of Commerce. Racial designations are defined as follows:

- **White**: A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

- **Black**: A person having origins in any of the black racial groups of Africa.

- **American Indian or Alaskan Native (AIAN)**: A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.

- **Asian or Pacific Islander**: A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.

UCR data does not include a category for Latino or other ethnicity measures.

Standardization of racial definitions does not guarantee all agencies determine or report race in the same way. Bias in the measures presented here will arise if the categorization of race differs in a way that is related to other factors influencing the commission of crimes. There is some evidence that this is the case. For instance, Krosch and Amodio (2014) find economic scarcity increases the likelihood that a mixed-race individual is perceived as Black.

---

[5] For a detailed discussion of the UCR methodology, see the Uniform Crime Reporting Handbook (U.S. Department of Justice and Federal Bureau of Investigation 2004) and FBI Data Quality Guidelines (Federal Bureau of Investigation 2018).

Offense Categories

UCR organizes arrests into 49 categories. See Uniform Crime Reporting Handbook (U.S. Department of Justice and Federal Bureau of Investigation 2004) for full definitions of original UCR categories.   Sample size constraints required collapsing these original UCR offense categories into the following broad categories:

1. **Violent Crimes**: Murder and non-negligent manslaughter; forcible rape; robbery; aggravated assault; other assaults.

2. **Theft or Property Taking**: Burglary-breaking or entering; larceny-theft (not motor vehicles); motor vehicle theft.

3. **Destruction of Property**: Arson; vandalism.

4. **Fraud or Dishonesty**: Forgery and counterfeiting; fraud; embezzlement; stolen property-buy, receive, and possession.

5. **Drug Sale or Manufacturing**: Opium, coke, and their derivatives; marijuana; truly addicting synthetic narcotics; other dangerous non-narc drugs.

6. **Drug Possession**: Opium, coke, and their derivatives; marijuana; truly addicting synthetic narcotics; other dangerous non-narcotic drugs.

7. **Total Drug Abuse Violations**: Combines "drug sale or manufacturing" and "drug possession" categories.

8. **Self-Directed Conduct Offenses**: Bookmaking (horse and sports); number and lottery; all other gambling (original ucr category); weapons (carry, posses, etc.); prostitution and criminal vice.

9. **Status Offenses**: Liquor laws; drunkenness; disorderly conduct; vagrancy. for juveniles, this category also includes juvenile-specific status offenses: curfew and loitering violations; runaways.

10. **All Other Non-Traffic Offenses:** Sex offenses (not rape or prostitution); offenses against family and children; driving under the influence; suspicion; all other non-traffic offenses (includes the original ucr category).

Agency Non-Response

UCR has been criticized for incomplete reporting by police agencies (Maltz and Targonski 2002).   Non-reporting becomes problematic when constructing measures of arrest at larger geographic levels because areas with significant amounts of missing data result in mass imputations at the national, state and county level (Lott and Whitley 2003, Maltz 1999).

To address this issue, we construct an *agency-level* measure of racial disparity in arrests. In contrast to larger geographic estimates, agency-level measures requires less imputation.

Federal agencies consider UCR data in the allocation of law enforcement funding, encouraging participation. Nevertheless, UCR is voluntary and on average 77.10 percent of agencies report in a given year, 19.12 percent of agencies do not report at all, and 3.77 percent report fewer than four months (considered non-reporting, and omitted from this analysis).[6] Reporting agencies should be regarded as a non-random subset of all agencies. If non-reporting agencies differ significantly from reporting agencies, this will bias summary statistics. This is discussed in depth below.

Indeed, agencies that are county-level and task-force agencies with special jurisdictions were less likely to respond, as were smaller rural agencies. Agencies were also less likely to respond if they had a previous-year arrest rate that was lower than the agency's overall average rate. Response rates increase over the course of the data period.[7] Our data contains at least one year of reported arrests from 10,599 municipal and township agencies; 2,866 county police and sheriff's offices; and 135 tribal authorities (see table 1).

[Table 1. Response Rate by Weighting Strata.]

Because state and federal agencies appear in the data sporadically these agencies are excluded from the data. We also exclude special jurisdiction agencies such as transportation authorities, parks, and recreation agencies, because it can be very difficult to determine the jurisdictional population these agencies regularly police. The final sample is limited to county, local, and tribal agencies. County agencies include both sheriff's offices and county police; local agencies include both municipal and township policing agencies.

Monthly Non-Response

Agencies that provide some monthly data in a year may still fail to report in all 12 months. Agencies that report at least four months in a calendar year, but have missing months are imputed through one of three methods: (1) nearest month imputation; (2) seasonally adjusted annual average; and (3) annual average, in preferential order. Across all years, 18.45% of agency-months are imputed.

(1) If the missing month is bracketed on both sides by a reporting month, the missing month is allocated as an average of the two nearest month's numbers. Thus, an agency that fails to report in April, but reports in both March and May, will have their April months arrests imputed as an average of March and May. Overall, 8.90% of agency-months (representing 48.25% of missing agency-months) are imputed using this method.

(2) Most studies which utilize UCR data impute missing months using an average of an agency's reported months in a given year. However, there is a known seasonal effect on crime (McDowall, Loftin and Pate 2012). This is true of arrests as well. Figure 1a displays the pattern

---

[6] A few agencies report their full year arrest count once a year, typically in December. In 2015, this was 384 agencies (0.22% of agencies). These agencies *are* included in the analysis.

[7] Tribal agencies have a very low response rate in all years except 2015 (see figure 2).

of monthly arrests between 1999 and 2015 for all reported arrests. The top of the figure displays the seasonal effect on total (non-imputed) arrests over the course of the year, with January functioning as the baseline, after controlling for agency arrest means. The form of the model is:

$$T_{amy} = c + \boldsymbol{a}\boldsymbol{M}_m + \boldsymbol{b}\boldsymbol{Y}_y + d\bar{T}_a + e \tag{1}$$

where T_{amy} is the total number of arrests made by agency a in month m and year y. Of primary interest is $\boldsymbol{a}'$, a vector of coefficients that estimates the effect of $\boldsymbol{M}_m$, a vector of indicator variables representing each month. The model also contains $\boldsymbol{Y}_y$, a vector of dummy variables for each year, and $\bar{T}_a$, the mean number of arrests made by the agency a during the reporting period. Consistent with previous research (Andresen and Malleson 2013), arrests peak in the warmer months and trail off at the end of the year as agencies grow tired of reporting.

We define seasonally reporting agencies as those that fail to report for three or more consecutive months as this creates a seasonal pattern of missing data. Figure 1b displays the average seasonal effect as calculated in equation 1, limited to seasonal reporting agencies. These agencies have a strikingly different pattern of arrests arising from the nature of their reporting.

We impute missing months for these agencies using $\boldsymbol{a}'$, the vector of coefficients that estimates the effect of arrests on month m. For the imputation, different seasonal effects are calculated separately for each offense, jurisdiction type (county, local, and tribal), and race. The coefficient on the indicator variable measures the extent to which arrests in that month tend to be higher or lower than agency's average. We then impute missing months by adjusting the month average of that agency's arrests that year by the monthly effect. Overall, 1.35% of months (representing 7.30% of missing months) are imputed using this method.

[Figure 1. Seasonality effect.]

(3) As a final imputation method, agencies that are missing no more than four months non-consecutively throughout the year are imputed using the average of all reported months in that calendar year for that agency, offense, and race group. Overall, 8.20% of months (representing 44.44% of missing months) are imputed using this method.

Table 2 provides descriptive statistics by imputation method. For local agencies, 15.8 percent of years have at least one month imputed using the nearest-month method, 1.3 percent are have a month imputed with a seasonally adjusted estimate, and 5.8 percent are imputed using the monthly average in the year. This is fairly consistent across agencies, with the exception of tribal police, which have substantially fewer months imputed using nearest month than other agencies.

[Table 2. Agency Descriptive Statistics, by Monthly Arrest Imputation Type.]

A logit regression predicting the likelihood that an agency-month is imputed, using any method, shows that local-level agencies are more likely to be imputed than county-level agencies. Agencies are also less likely to be imputed if they have a high population density or a large jurisdictional population. Results of the regression are displayed in table 3.

[Table 3. Logistic Regression Modeling Odds of Imputed Monthly Arrest Estimates.]

Non-Response Post-Stratification Weights

Agencies which report fewer than four months of the year are considered to be non-responding and are omitted from the data. The remaining agencies are a non-representative subset of all law enforcement agencies nationally. For larger geographies or yearly trends analysis, this non-representativeness is problematic. For instance, if annual trends show variation, it can be difficult to determine if this variation derives from changes in in actual arresting behavior, or simply variance resulting from changes in the number and type of agencies reporting changes.

We construct post-stratification design weights to partially correct for agency non-response where weight for strata H is calculated as:

$$W_a = \rho / \widehat{\Phi}_H \tag{2}$$

where

$$\widehat{\Phi}_H = \sum_{a \in R_H} \pi_k^{-1} \Big/ \sum_{a \in P_H} \pi_k^{-1} \tag{3}$$

Where, π_k^{-1} is the inverse probability of selection, ρ is the size of the total arrestable population in the agency's jurisdiction, R_H is the set of agencies from strata H that responded and P_H is the set of agencies sampled that responded. Because UCR data comes from a voluntary reporting of all agencies nationwide, there is no sampling design affecting probability of selection and $\pi_k^{-1} = 1$.[8] Therefore equation 3 simplifies to:

$$\widehat{\Phi}_H = n_H / N_H \tag{4}$$

or the response rate for strata *H*.

To construct post-stratification weights, it is necessary to determine the number of agencies that could have responded in a given year. This requires a complete census of all law enforcement agencies. An agency meets the federal definition of a criminal justice agency when it is "a governmental agency or any subunit thereof that performs the administration of criminal justice pursuant to a statute or executive order, and which allocates a substantial part of its annual budget to the administration of criminal justice" (Title 28, Code of Federal Regulations [CFR], Part 20, Subpart A).

We construct a list of agencies utilizing both the Census of State and Local Law Enforcement Agencies (CSLLEA), available from the Bureau of Justice Statistics, and the Law Enforcement Agency Identifiers Crosswalk (LEAIC), available through the Inter-University Consortium for Political and Social Research (ICPSR). These contain overlapping, but not identical, lists of agencies, as well as unique agency identifiers (Originating Agency Identifier or ORI). We

---

[8] For a discussion of weighting with non-probability data, see (Little and Vartivarian 2003).

combine these agencies with the list of agencies that ever reported to the UCR program from 1999 to 2015. Each list represents a partially complete subset of law enforcement agencies in a given year.

Because the definition of agency may include special task-force operations, or other sub-agency function created for purposes as needed, agencies may not exist in all years or consecutively throughout of the data. When an agency is missing it can be difficult to determine if that agency did not exist or is simply not reporting. Therefore, we assume an agency is 'born' in the first year it appears in the combined list, and an agency 'dies' the final year it is included in any list source. The agency is then assumed to exist during all intermediate years. Our data includes only municipal, county, and tribal jurisdictions, and excludes special jurisdiction agencies that do not fall into one of these three jurisdictions. We do include special task force agencies if those agencies are under the jurisdiction of an included agency. For example, the drug task force for a local police agency would be included in the sample. With the exception of special jurisdictional units, agencies are assumed to have the same type of government and jurisdiction over the full timeframe.

Our data contains at least one year of reported arrests from 10,599 (96.3%) of the 11,009 municipal and township agencies; 2,866 (98.6%) of the 2,907 county police and sheriff's offices; and 135 (90.0%) of the 150 tribal authorities (see table 1). Of the total 14,067 agencies listed 13,547 (96.3%), enter the data frame in 1999 and all but one exits the data frame in 2015.

[Table 4. Year of Birth and Death for All Agencies and Responding Agencies.]

In constructing weights, agencies were stratified by area jurisdiction type (county, municipal, or tribal) and the agency's jurisdictional population density (divided into three categories). These categories are constructed using one of the many federal definitions of rural and urban.[9] Areas with fewer than 1,000 people per square mile are classified as non-urban. We further subdivide the non-urban classification into areas with fewer than 250 people per square mile. Agencies may move between strata in different years if their jurisdiction or population density changes. Weights are not trimmed. Total population weights have been rescaled to sum to the U.S. population in a given year. Racial group population weights have been rescaled to sum to the U.S. group population in a given year.

Figure 2 presents response rate by strata. The number of agencies reporting increases slightly over the time span. In 1999, about 26.3% (3,567) of agencies fail to report at all and 12.9% (1,742) report less than four months of data. By 2012, the percent of agencies not reporting declined to

---

[9] This is the definition used by Department of Health and Human Services, as well as the Department of Veteran Affairs, to classify health care providers. For more information see https://www.nal.usda.gov/ric/what-is-rural.

13.5% (1,904), but the number of agencies reporting 1 to 3 months of data increased to 22.3% (3,135).[10]

[Figure 2. Response rate by weighting strata.]

The dataset include three weights. The first is a base weight that adjusts for non-response, but is not adjusted for agency's jurisdictional population. In these weights $\rho=1$. These weights should be used in analyses that make conclusions about agency behavior. The second weight variable is adjusted by jurisdictional population, and scaled to total U.S. population. These weights should be used when estimating the arrest experience of the 'average' American. The final set of weights is adjusted by jurisdictional population by race group, scaled to the total U.S. population. These weights should be used when estimating the arrest experience of the 'average' U.S. resident of color.

Population adjusted base weights increase the standard-error of estimates. The bias reductions gained by using these weights can easily be eliminated by this larger variability, making such weights non-informative. To determine if these weights are informative, we use a test pioneered by Pfeffermann and Sverchkov (1999).

The set of responding agencies is a subset of the population of agencies $(R \subseteq P)$. An unweighted mean of arrests for all agencies can be modeled [11] by:

$$Y_a = \mu + \varepsilon_a \tag{5}$$

and the weighted mean of arrests for all agencies can be modeled by:

$$Y_{w,a} = w_a \mu + \varepsilon_{w,a} \tag{6}$$

Pfeffermann and Sverchkov (1999) argue that, if the sample distribution of the residuals is the same as the population distribution of the errors, then one can ignore the sampling scheme and apply the unweighted regression. Their test is grounded in the idea that by comparing the distributions of the sample and population residuals, one can determine the contribution of weights to estimates. They suggest that $E_p(\varepsilon_i^k) = E_R(\varepsilon_i^k), k = 1,2,\dots$, where E_p and E_R are the expectations under the population and the responding probability density functions, respectively. Their paper also derives that $E_p(\varepsilon_i^k) = E_R(\varepsilon_i^k) / E_R(w_a)$.

$$\boldsymbol{W} = \boldsymbol{\alpha}_w + \beta\hat{\boldsymbol{\varepsilon}}_w^k + \boldsymbol{e}_w \tag{7}$$

After regressing the weighted residuals on the post-stratification weight variable, a t-test on the weight coefficients will estimate the relationship between the weights and the unexplained

---

[10] The response rate for tribal agencies is very low prior to 2015. We strongly recommend against conducting analyses with these years. Researchers should use extreme caution when comparing tribal agencies across time.

[11] The responding subset is not a random subset and thus $\hat{Y}$ is not an unbiased estimator.

variance in the model. Rejection of the null hypothesis, $H_0: \beta = 0$, indicates that sample weights are informative. For a discussion see Bollen et al. (2016). Our test indicates that all population adjusted base weight are informative (table 5).

[Table 5. Descriptive and informative test statistics for weighting variables..]

DETERMINING JURISDICTIONAL POPULATION

Because racial groups are not distributed uniformly across police jurisdictions, a simple count of the number of arrests by race is insufficient to understand racial disparity. For instance, an agency that arrests 10,000 Whites and 30,000 African-Americans may be exhibiting significant bias if the arrestable population is 70 percent White. Yet this agency may be demonstrating no bias if the arrestable population is 25 percent White. Therefore it is necessary to adjust arrest numbers to take into account the population at risk of being arrested by an agency.

This requires assumptions about the population the agency polices. In general, we assume that individuals living within the agency's jurisdictional geographic boundaries are the individuals most likely to be arrested by that agency. This means individuals traveling to the area for reasons such as leisure and work are generally excluded from the agency's at risk population, even though the agency may arrest these individuals if they commit a crime while within the jurisdiction. There is very little research supporting this assumption.[12]

POPULATION DATA

Geographic population estimates come from the U.S. Census and the American Community Survey (ACS). The ACS uses a series of monthly samples to produce estimates of small geographic areas. For very small geographic areas, five-year samples are needed to produce reliable estimates. This means that data from 2015 includes the years 2013-2017 and 2014 includes years 2012-2016. Larger geographic areas are estimated utilizing 3-year samples. As a result of these rolling estimates, population fluctuations are smoothed. It also means demographic changes are somewhat lagged. As data are not available in all years, missing years are linearly interpolated.

The ACS excludes institutionalized populations. However, one of the benefits of incorporating yearly data is that, to whatever extent arrestees are removed from the population though incarceration, this removal is reflected in the data. Whether the multi-year lag is problematic for the removal of arrestees is unclear, as incarceration also lags behind time of arrest.

Area of Likely Arrest

Though it contains no data on population by race, UCR contains a reasonable estimation of the number of people living within the jurisdiction of *municipal* agencies. However, to estimate the population living within the jurisdiction of a county agency, UCR subtracts the population of each participating local law enforcement agency within the county from the total county population. This means that sheriff and county police arrests are assumed to be limited only to individuals living in non-incorporated areas. In non-rural counties where most of the population lives in cities

---

[12] A study on tendency for arrests to occur in the same area or neighborhood of offender residence finds that drunk driving arrests tend to occur close to home (Lapham et al. 1998).

large enough to have a police force, the result is a very small jurisdictional population and drastically inflated arrest rates.

This is problematic because, in practice, most county agencies have arrest authority anywhere in the county (Burch 2007). It is generally unreasonable to assume a person seen committing a crime on a city street will not be arrested by a county officer, simply because that person could also have been arrested by city police. Some county agencies may choose to prioritize the policing of unincorporated areas because these areas are uncovered, but this is an agency-level discretionary decision.

In some instances, police may even have arrest authority outside their jurisdiction, depending on the nature of the crime and inter-government arrangements and agreements (Reaves 2015). Some cities even have contractual arrangements with county-level agencies to do some or all of their policing activities (Burch 2007). In these instances, the arrest is typically attributed by UCR to the county agency, even though it took place within a municipal jurisdiction.

Finally, some agencies choose to pool authority. This is most commonly seen in special jurisdiction task forces where multiple agencies police all their inclusive jurisdictions for specific activity, such as drug sales or counterfeiting. Because special jurisdiction task forces are examples of multiple agencies working in tandem across multiple jurisdictions, the appropriate population at risk of arrest is difficult to determine. These agencies have been dropped from the data.

Estimating Jurisdictional Population

We estimate jurisdiction using ACS data using the following procedure:

1. For county agencies, we use Census data from 2000 and 3-year ACS estimates from 2007-2015. Missing years are interpolated linearly. The jurisdictional population is assumed to be the same as the county population

2. For municipal, city, and township agencies, we use Census data from 2000 and 5-year ACS place and zip code estimates from 2009-2015. Missing years are interpolated linearly.

   a. Agencies are matched to Census places utilizing a fuzzy match algorithm using name of place, name of agency, and state. Any agencies with a match less than 99 percent (based on minimum edit distance) was manually verified.

   b. Agencies were matched to zip codes using LEAIC crosswalks, which lists the zip code range covered by the agency. Agency jurisdictions are assumed to be all zip codes within the range. This may pose a problem for agencies with multiple zip codes in the range if the zip codes within that range are non-contiguous. Approximately 5.03 percent of municipal agencies have more than two zip codes in range.

c. Local agencies are assigned either place or zip code data, depending on which area best matched the agency-level population estimates provided by UCR.

3. For tribal agencies, we use Census data from 2000 and 3-year ACS estimates from 2007-2015. Missing years are interpolated linearly. The presence of allotment and checkerboarded reservations, coupled with issues surrounding the Major Crimes Act (18 U.S.C. § 1153) and cross-tribal differences in the interaction of tribal and state law, mean that tribal agencies can differ significantly on policing behavior, making jurisdiction very difficult to predict. In the interest of simplification, it is assumed tribal authorities are most likely to arrest individuals residing anywhere within the county.

Finally, some special jurisdictions have indeterminate jurisdictional populations. These include places where people tend to live, but the demographics of the resident population is difficult to determine (i.e. college campuses, military bases, etc.). Additionally, some police agencies have jurisdictions (and make arrests) where no individuals reside (i.e. airports, state parks, etc.). Because such agencies tend to specialize or see a very specific subset of crimes, and the nature of their policing likely differs from 'full-service' agencies, we exclude both forms of special jurisdictions from the analysis.

The final set of agencies contains 14,067 agencies, including 2,908 county agencies, 11,009 municipal police, and 150 tribal police.

Construction of Arrest Rates

Population numbers are used to construct arrest rates. Arrest rates are calculated as:

$$AR_{ato}^r = \frac{T_{ato}^r + \Pi_{at}^r}{P_{at}^r + 1} \tag{8}$$

where AR_{ato}^r the arrest rate for race r for agency a and offense o in year t. T is the total arrests for the year, P is the agency's jurisdictional population, and Π is the proportion of the agency's jurisdiction that is race r.

It is possible for agencies to make no arrests of a racial group for a specific offense in any given year. This is most likely to occur in areas with very small minority populations. This zero numerator will become problematic for later analyses. We therefore apply a population adjustment to the numerator where the number of arrests is increased by the proportion of the population in the agency's jurisdiction represented by that racial group. For example, an agency that arrested two African-Americans in a given year and has a jurisdiction that is 5% Black will have a numerator of 2.05 rather than 2. This adjustment has the greatest effect on areas where the racial population is small and the agency tends to make very few arrests.

The adjustment deflates arrest rates for a small subset of agencies. Most of this reduction results from the additional person in the population (+1 in the denominator), meaning that most of the difference results in agencies where there are no residents of the minority population. Thus, the additional person in the denominator makes the assumption that, even if no person of that race

lives in the agency's jurisdiction, at least one person will pass through their jurisdiction during the calendar year. This is possibly an unreasonable assumption in some small, isolated locations.

[Figure 3. Arrest rates with and without numerator adjustment, by race.]

Despite the procedure matching local agencies to their jurisdictional population, described in (c) above, some jurisdictional mismatch may still occur to the extent an agency polices a high traffic area. For instance, a small municipality that includes the county's fairground or event center would see a significant increase in non-town visitors arriving from the surrounding county during events held at the fairground. If such events are frequent enough, this would change the population the agency is likely to regularly police.

Unfortunately, such agencies are difficult to identify without anecdotal local information. We identify such agencies using outlier analysis. If an agency reports an arrest rate that is an outlier for more than 50 percent of their years in sample, and they reported for more than 3 years in the time frame we assume the high arrest rates are generated by jurisdictional mismatch and not by an underlying policing practice. An outlier is defined as having an arrest rate for all races combined that is greater than one and is also greater than 95 percent of all other agencies of that jurisdictional type in that year. However, the outlier procedure never identifies an agency with a rate less than 1. A rate of one would mean the agency arrested every person in their jurisdiction once. This seems an unlikely possibility.

Mismatch identified agencies have their jurisdiction adjusted to the next largest population level. Consider the example of an agency that has a place population size of 3,000; a zip code population of 5,000; and a county population of 60,000. If the agency was originally assigned a place population and was tagged as a jurisdictional mismatch, the agency would be assigned zip code level population data and arrest rates would be recalculated. If the agency again indicates mismatch, the agency would be assigned county-level population data. If the agency still exhibits outlier properties, these are assumed to be a real and correct observation, and the agency's jurisdictional population remains at county-level. We also perform this jurisdictional correction on any agency with fewer than 75 adults in the area jurisdictional population (364 agencies). Jurisdictional mismatch corrections are performed for local agencies only; county and tribal agencies are excluded.

The number of agencies identified as outliers using this procedure was small. Overall, 23,140 agency-years (12.76% percent of all agency-years) experience an outlier across one year or more for at least one race-age group. Table 6 displays the number of outliers identified by agency jurisdiction. Most outliers tend to co-occur across demographic groups. In other words, an agency with an outlier on one race also tends to have outliers across the arrest rates of other races (including White). This suggests being an outlier by this definition is indeed a problem with the population denominator and not simply a sign of racial bias.

[Table 6. Number of agency-years with identified outliers.]

Overall, 43.81 percent of outliers occur only once or twice per agency. These rare outliers are not addressed using population adjustments. Only 14.84 percent of agencies with outliers have outliers that occur on more than 50 percent of their observations. Overall, a total of 734 agencies (5.2% of all agencies) required a population adjustment. The iterative assignment of areas is displayed in Table 7. Most outlier adjustments occur in small, rural, municipal agencies, particularly when the population of analysis is small.

[Table 7. Iterative assigned population jurisdiction, by agency type.]

Correcting arrest rates in this manner assumes that consistently high arrest rates are a data artifact rather than a real observation. This is likely not always the case, as some agencies may simply make unexpectedly large numbers of arrests. The population adjustment discussed here creates conservative arrests rates, meaning we assume high arrests are unreasonable and actively seek to reduce rates. Because outliers tend to occur within an agency across racial categories, the editing procedure is unlikely to substantially impact measures of racial bias.

MEASURING RACIAL DISPARITY IN ARREST

FINAL SAMPLE

The final set of agencies contains 14,067 agencies (1,994,938 agency-years), including 2,908 county agencies (452,254 agency-years), 11,009 municipal police (1,539,494 agency-years), and 150 tribal police (3,190 agency-years). Data contain information on 172,474,735 arrests over the 17 year period. This is enough arrests for every adult, age 18-64, in the U.S.in 2015 to have been arrested once. Overall, the data contains 118,345,835 White arrests, 49,573,475 Black arrests, 2,076,087 Asian arrests, and 2,490,536 AIAN arrests. Table 8 displays descriptive statistics for the full period, by agency type.

[Table 9. Descriptive statistics for responding agencies, 1999-2015, by agency type..]

MEASURE 1: MEASURING A RACIAL GAP IN ARREST RATES

The first measure we develop is a simple risk ratio that takes the demographic composition of the arrestable population into account, defined as

$$RR_{ato}^r = ln \frac{AR_{ato}^r}{AR_{ato}^{White}} \tag{9}$$

where the numerator is the arrest rate for a non-white population (Black, Asian, or AIAN, measured separately) and the denominator is the arrest rate for the White population, where a denotes the agency, t denotes the year of arrests, and o denotes the offense.[13]

Figure 4 displays the distribution of logged risk ratios for each demographic population. There is a non-trivial number of agencies that arrest no individuals of a race-age group in a specific year. For instance, about 90 percent of agencies arrest no AIAN juveniles and about 85 percent arrest no Asian juveniles. Arrests for sub-categories of offenses are even more likely to be zero, as more than 95 percent of agencies make no AIAN arrests for violent crime. The inflated number of zeros may be problematic, depending on the type of analysis, and we urge researchers to use appropriate methods. The adjustment of these zero observations is discussed in equation 8.

[Figure 4. Distribution of risk ratio for adult and juvenile arrests, by race]

On average, there are 8.5 Black adult arrests, 1.1 Asian adult arrests, and 2.4 AIAN adult arrests for every one White adult arrest.[14] However, looking only at places that make at least one arrest

---

[13] Without taking the natural log, risk ratios are highly skewed and show a high degree of kurtosis (for example see LaFree, Baumer and O'Brien 2010). The sampling distribution for the natural log of a risk ratio is approximately normal (for a proof see Fleiss 1993).

[14] Comparing risk ratios between demographic groups should be done with caution, as differential identity issues will confound comparisons. This is most readily visible in estimates of AIAN arrests, as identification of Natives in the Census and ACS is notoriously noisy (Eschbach, Supple and Snipp 1998) and may tend to overestimate the AIAN

among the population in question, these numbers become 15.8, 5.4, and 16.5, respectively.[15] Summary statistics for risk ratios are presented in table 9.

[Table 9. Summary statistics for logged risk-ratio, by agency and race.]

A logged risk ratio provides an easily interpreted measure of relative arrests. A logged risk ratio greater than zero means that the non-white group has a higher risk of arrest than Whites. A risk ratio less than zero means whites have a higher arrest risk than the non-white group. It is an excellent summary of the policing experience of a specific group. However, since the risk ratio does not take into account racial differences in the underlying rate of commission of crime, it does not provide a measure of agency behavior. Therefore, it is best used when researchers have no need to distinguish between crime rates and policing behavior.

It is important to note that, while the risk ratio provides a measure of relative risk, it requires context to be appropriately interpreted. For instance, when the risk ratio is close to zero, yet the arrest rate for the non-white group (denominator) is high, the risk of arrest affects a large number of people. A logged risk ratio near zero simply means the risk is balanced between groups but does not speak to the magnitude of social significance.

The risk ratio also assumes the correct comparison group when evaluating racial disparities is Whites. In essence, we are assuming the White rate represents a 'just' and 'fair' rate.[16] There are two flaws with this assumption. First, several commentators have noted that the assumption of a White comparison group is itself a form of bias (for an example, see Allport 1954 (1979)). Additionally, UCR contains no category for Latino/Hispanic arrests. As such, it is likely the bulk of these arrests are included in the 'White' category. If there is any racial bias in arrests of Latinos, this will artificially increase the denominator, resulting in conservative risk ratios. Results should be interpreted judiciously.

Figure 5 displays trends for risk ratios with both base-weights and population-adjusted base weights. Between 1999 and 2015, racial disparity, measured by the logged risk ratio, increased across all three racial groups.

[Figure 5. Risk ratio for adult and juvenile arrests, by race]

---

population. Simultaneously, third party identification used in arrest numbers likely underestimates the number of AIAN arrests. This will produce a deflated arrest rate and risk ratio, thus making cross-racial comparisons difficult.

[15] Researchers should note when using data that smaller areas, agencies, and offenses produce higher variance and create less stabilized estimates.

[16] This ignores questions of overcriminalization, and whether some actions are crimes that should not be crimes.

Outlier Analysis

Some outliers continue to occur. These are most common in small jurisdictions with very small minority populations. For example, a small municipality with only one African-American resident that arrests three African-Americans passing through the jurisdiction will have a high arrest rate. After adjusting the arrest rate for population mismatch (see above) we believe these represent real arrests. The potential of these outliers to influence results depends on the nature of the analysis being conducted by the researcher. Figure 6 displays trends in the logged risk-ratio with outliers omitted. In general, while point estimates are sensitive to outliers, trends are less so. We strongly suggest researchers conduct outlier analysis when using the data.

[Figure 6. Risk ratio for adult and juvenile arrests, estimated with full sample or with outliers.]

MEASURE 2: CORRECTING DIFFERENTIAL COMMISSION

Arrest rates are a combination of two processes: (1) law enforcement policies and procedures that result in arrest, and; (2) the actual rate at which crimes are committed (Hindelang 1981, O'Brien 1996, O'Brien 2003). The risk ratio presented above provides a good baseline measure of relative arrests, but fails to distinguish between these two processes. Therefore, it is important to distinguish between racial disparity and racial bias (Goff and Kahn 2012). Disparity is the simple act of arresting more of one group than another. However, bias, put crudely, is when an agency arrests a specific sub-group at a higher rate than it "should". This raises the question: "At what rate *should* an agency be arresting a group?" The deceptively simple answer is that an agency should be arresting individuals who commit crimes. However, the actual rate of crime commission is unobservable, in that is there is no actual measure of how many crimes are committed. As researchers, the best we can hope for is a measure of the number of crimes reported, observed, or solved. Before we can address whether an agency is arresting at a rate higher than it "should", we must first estimate the underlying rate of crime commission within that agency's jurisdiction. We do this by controlling for factors that have been empirically demonstrated to affect actual rates of criminal offending.

By regressing these explanatory variables, which correlate to the commission of crime on arrest rates, we explain, at least partially, the portion of variance in arrest rates resulting from the crime rate, leaving the variance attributable to police action unexplained.

These control variables relate to several theories of the causes of crime. These theories fall into two broad camps (for a detailed discussion see LaFree, Baumer and O'Brien 2010, Phillips 2002). The structural disadvantage camp argues that social pressures, such as unemployment, poverty, income inequality, or inferior education, disproportionately affect minority populations, leading to greater levels of frustration and aggression, and ultimately resulting in higher crime rates (e.g. Blau and Blau 1982, Liska and Chamlin 1984, Messner and Golden 1992, Miethe, Hughes and McDowall 1991, Steffensmeier and Dana 2000). A second narrative emphasizes cultural differences across demographic groups, stressing that different historical experiences of minority populations create a different value system that ultimately affects the likelihood of

committing crime (e.g. Miethe, Hughes and McDowall 1991, Neapolitan 1994, Nivette 2011). These narratives are not incompatible with one another in that, if cultural distinctions exist, they likely vary with features of structural disadvantage (see Sampson and Wilson 1995). Neither of these theories suggest there are racial differences in being individually predisposed to commit crime.

Past research found numerous factors associated with crime. These factors are listed in table 10. Each study measures these variables in different ways. While we would prefer to incorporate as many factors as possible, to explain as much variance as possible, many are highly correlated. When this occurs, we have chosen to use the measure that explains the most variation. The final set of control variables includes:

- Nonwhite-White median income ratio; [17]
- Nonwhite-White poverty rate ratio;
- Number of female-headed households;
- Rate of between-county geographic mobility;
- Nonwhite-White high school dropout rate ratio;
- Nonwhite-White educational attainment ratio for less than high school;
- Nonwhite-White educational attainment ratio for high school completion;
- Nonwhite-White unemployment rate ratio;
- Population density;
- Population size;
- Overall arrest rate;
- Proportion of the population under age 18;
- Hispanic population proportion;
- Housing rental rate; and,
- Housing vacancy rate.

We also examined the Nonwhite-White ratio of male earnings, gini coefficient, proportion of children in two parent households, Nonwhite-White divorce rate ratio, female fertility rate for ages 15-19, Nonwhite-White male educational attainment ratio, Nonwhite-White labor force participation rate ratio, proportion of families in multi-generational households, and the probability that randomly selected dyads are a Nonwhite-White dyad. These variables were dropped due to multicollinearity and lack of additional explanatory power.

[Table 10. Variables empirically demonstrated to correlate with crime rates.]

It is important to note some researchers found that the processes which generate crime differ in their effect by racial group (Harer and Steffensmeier 1992, Krivo and Peterson 1996, LaFree,

---

[17] All Nonwhite variables are measured separately as Black, Asian, or AIAN, depending on the demographic group under analysis.

Drass and O'Day 1992, Messner and Golden 1992, Sampson 1987, Shihadeh and Steffensmeier 1994). Descriptive statistics for these control variables are displayed in table 11.

[Table 11. Descriptive statistics for residual regression control variables, full sample.]

Because the population of law enforcement agencies is small, we model crime commission using the finite population regression model, the form of which is:

$$RR_N = X_N B + E_N \tag{10}$$

with residuals defined by:

$$E_N = RR_N - X_N B \tag{11}$$

where RR_N is a N × 1 vector, X_N is a N × k matrix of explanatory variables with the first column of ones for the regression constants, B is the k × 1 corresponding vector of coefficients, and E_N is an N × 1 vector of errors.

By construction, E_N has the following properties:

$$1' E_N = 0 \tag{12}$$

and

$$X_N' E_N = 0 \tag{13}$$

That is, all of the omitted influences have a mean of zero when combined, the variance of the error is the same for all cases (homoscedasiticity), and the errors of different cases are uncorrelated (no autocorrelation). The model also assumes the error variable is distributed independently of the explanatory variables.

For estimability, it is also required that:

$$rank(X_N) = k \tag{14}$$

In other words, if the explanatory variables are highly correlated, then $X_N' X_N$ has no inverse.

Of interest is E_N, the residual, which measures the variation in arrest risk-ratios unexplained by the underlying commission of crime. Although, because the model is an imperfect estimation of the ratio of crime commission, it is better to think of E_N resulting from law enforcement behavior and an error term. Agencies with a positive residual ($E_N > 0$) are arresting a demographic group at a rate higher than expected.

Estimation is done on a reduced sample that excludes observations with an absolute value studentized residual greater than 3.5. We also exclude observations with a leverage higher than (2k+2)/n, where k is the number of predictors (here k=15), and n is the number of observations. Finally, we exclude any observation with a Cook's D greater than 4/n. The goal is to estimate the effect of each factor on crime commission without allowing agencies which are displaying outlier arresting behavior to unduly influence estimates. However, because we believe these outliers

represent real-world policing behavior, we use the estimated $\boldsymbol{B}$'s, in equation 11, and estimate $\boldsymbol{E}_N$ for the full sample, including omitted outliers. Because of this $\overline{\boldsymbol{E}}_N \neq 0$.

Model Fit

Because it is plausible the processes generating crime differ across groups and may change over time, the residual regression model is run separately on each year (1999-2015), each of 11 offense categories, and across 3 racial groups and 2 age groups, resulting in 1,122 separate models.

Table 12 summarizes several measures of model fit for the residual models. Because many of these factors measure disadvantage, multicollinearity between the explanatory variables is a concern. Therefore, we also measure variance inflation for each predictor.

In the analysis of geographic areas, heteroscedasticity is a potential problem, because the error of variance declines as population size increases. We test heteroscedasticity by population size using the Breusch-Pagan test (Greene 1993: 394-395). Analysis indicated some models have significant heteroscedasticity. Heteroscedasticity does not bias coefficient estimates, but does lower precision.

[Table 12. Descriptive statistics for regression diagnostics from 1,122 residual models.]

Measure of Disparity

Unlike the risk ratio, the resulting residual measure of disparity does not have a straight forward numeric interpretation. In general, positive numbers indicate an agency arrests more of a specific demographic group than would be expected, given observed factors correlated with the commission of crime, and a negative number indicates fewer arrests than would be expected. Summary statistics for the residual measure are located in table 13. Across both adult and juvenile arrests, the residual for Black arrests is 0.020, for Asian arrests is 0.004, and for AIAN arrests is 0.007. If we include only agencies that arrest at least one person from the demographic group, the numbers increase to 1.05, 2.19, and 3.40 respectively.

[Table 13. Summary statistics for regression residual, by agency and race.]

On average, individual agencies experience an average increase of about 0.001 in their residual from the prior year. As the standard deviation in the residual is 0.83, this is a minuscule change from year to year, indicating the measure is relatively stable for most agencies.

The residual is a valid measure of racial disparity only to the extent that it nets out the actual commission of crime. Since this is unknowable, the measure is sensitive to omitted variable bias. One of the benefits of the using the risk-ratio is that the ratio structure reduces, but does not fully eliminate, concerns of omitted variables. By comparing factors related to underlying propensity between two groups (the group of interest and Whites), we reduce the burden on the model to completely predict differential propensity. Rather, the model is biased only to the extent these factors generate crime differently across race. While this is still clearly a strong assumption, it improves upon a basic control model.

The propensity-controlled residual measure is more directly comparable across time and place than the risk ratio. However, because the measure relies on demographic and socio-economic variables that correlate with crime rates, it is not possible to then study how these *same* variables effect racial bias in arrests. In other words, to create a more valid measure of racial bias in arrest, some analytical leverage has been traded away.

MEASURE 3: MAKING COMPARISONS ACROSS AGENCIES AND TIME

The definitions of laws differ across states and over time. The elements of a crime, even basic seemingly universal crimes, depend on legislative idiosyncrasy. Additionally, agencies at different jurisdictional levels often police or prioritize different types of crime. Because of this, comparison of the residual measure across states, agency types, and over time can become problematic.

To correct for these structural differences, we standardize the residual by the type of offense, state, and year. All cells are standardized to a mean of zero (0) and a standard deviation of one (1). In essence, the standardized residual compares an agency to other agencies that perform similar tasks and exist within the same legal framework. This provides two advantages. First, it allows for better comparisons between agencies. Agencies with a large positive standardized residual are arresting a demographic group at a higher rate than similar agencies, controlling for crime commission. The second benefit is that the standardized residual helps address remaining omitted variable bias still exhibited by the residual. To the extent an omitted variable exhibits racial variance, this invariance is likely similar for agencies that are in close geographic proximity. Nearby agencies act as a form of control on the residual, helping remove the remaining influence of omitted variables.

The standardized residual is very conservative. While it corrects for some remaining sources of variation that are not attributable to agency behavior, because it is constructed to average to zero across place and time, the standardized residual cannot be used to compare geographies that cross state lines, or used to make year-to-year comparisons. Again, we have traded analytic leverage for rigor. The measure is best used in examining agency-level behavior rather than trends.

VALIDATING MEASURES USING DOJ INVESTIGATION DATA

The current gold standard for assessing racial bias is the law enforcement misconduct investigation, conducted by the Department of Justice. After the enactment of 42 U.S.C. § 14141 allocated authority, the Violent Crime Control and Law Enforcement Act of the DOJ instigated 68 investigations across 54 agencies between 1999 and 2015. These investigations are one of the primary mechanisms for oversight and correction of "institutional failures that cause systemic police misconduct" (U.S. Department of Justice 2017). As these investigations arguably provide the only U.S. data on verified racial bias, we utilize information from these investigations as an external validation of the measures presented here.

Such investigations frequently take months or years to complete. Department investigators engage in interviews and outreach with area experts, and employees of the agency at all levels,

and conduct complex statistical analysis of documents collected on site. As an example, during the investigation into the Ferguson Police Department, investigators spent approximately 100 person-days on-site, which included: interviews of officials including City Manager, Mayor, Judges, Chief of Police; ride-along's with on-duty officers; and hundreds of interviews with residents and civic associations. Additionally, investigators reviewed 35,000 pages of department records.

We divide these cases into four types:

(1) investigations for which the DOJ decided not to pursue reform agreements or settlements (presumably because the agency behavior did not rise to a level consistent with the department's standards) – 17 agencies;

(2) agencies where the agency found problems unrelated to racial bias or discrimination – 24 agencies;

(3) agencies where the department found evidence of racial bias or discrimination against blacks and African-Americans – 5 agencies; and,

(4) agencies where the department found evidence of bias or discrimination against another group – 8 agencies.

Figure 7 summarizes these investigations.

[Figure 7. Summary of DOJ investigations.]

Because the DOJ does not have the resources to investigate all complaints of misconduct, priority is given to instances when the Assistant Attorney General believes: (1) the allegations, if true, establish a violation of the constitution or federal law; and (2) the allegations, if true, represent a pattern or practice, rather than sporadic isolated violations. Because this procedure identifies more agencies than might be investigated given current resources, the department prioritizes cases where "reform can have broad ranging effects or represent an emerging issue where federal action may help set a standard for reform" (U.S. Department of Justice 2017:6). This frequently includes unlawful stops and arrests, retaliatory force, and bias in race, gender or nationality.

The result of this selection process is that investigated agencies in no way resemble a representative sample of agencies. Indeed, they are representative of very unusual circumstances. It is therefore, not informative to compare measures of racial bias on the handful of agencies investigated by the DOJ to the population of all agencies. Instead, our analysis compares agencies where the Department found racial bias (types 3 and 4) versus agencies where it did *not* find racial bias (types 1 and 2) yet bias was alleged."

Figure 8 compares the measures developed in this paper between these highly specific samples: investigations with unverified allegations of bias, and investigations finding racial bias.

[Figure 8. Bias measures by DOJ findings.]

Across all three measures of bias against African-Americans, agencies with unsubstantiated allegations have lower than average levels of racial bias, compared to the full population of agencies (though, again, we caution against this comparison). We divide investigations finding racial bias into investigations finding racial bias against African-Americans, versus those finding bias against another population.[18] Investigations finding bias against African-Americans have higher than average scores of Black bias on all three of our measures, compared to investigations with unsubstantiated bias allegations. Investigations finding bias against a race that is not African-American do not consistently score higher on measures of Black bias, though it is difficult to say if agencies with such finding should, as it is unclear what the relationship is between non-Black bias DOJ findings and our measures of Black bias. We would *strongly* caution against the over-interpretation of these results, as the sample includes only five agencies, but the fact that these agencies score high on our measures of Black bias provides a nice sanity check.

Past research shows that studied agencies improved after implementation of a consent decree, though the sustainability of that improvement varies substantially (Bromwich Group LLC 2016, Chanin 2014, Davis, Henderson and Ortiz 2005, Stone, Foglesong and Cole 2009).

We create a difference-in-difference estimate of the three measures of racial bias using the following function:

$$DID = (R_t^B - R_0^B) - (\overline{R_t^N - R_0^N}) \tag{15}$$

where R is a measure of racial bias (the risk ratio, residual, or standardized residual), B denotes investigations where the DOJ found racial bias against Blacks, and N denotes investigations with allegations of bias but no finding. The estimator is constructed by subtracting the measure, R, in year t, from the year in which the investigation began, notated as year 0. In other words, DID estimates the change in bias, after the investigation, while simultaneously removing contemporaneous trends in bias from the control sample during the same year. We then average the DID estimator across the number of years before and after the start of the DOJ investigation. Because the estimator averages the control group prior to netting out the effect of the control trend, the DID estimator will not be zeroed in the start year.

Figure 9 depicts the trends in the DID estimator for the years surrounding the start of an investigation. The DID estimator does not include a confidence interval, as it is not a statistic estimating a population behavior but rather a measure taken on the entire population of investigations conducted during the time period. For all three measures of bias, the DID estimator is mostly stable in the four years prior to the investigation, declines slightly in the year the investigation begins, and then increases the following year before declining markedly. The decline in the years following investigation is consistent with past research, though we have insufficient

---

[18] In most instances, bias against another race is against Hispanics/Latinos, which are not measured in our data. In a few instances, the bias is found against a highly specific population such as young females or low-income whites.

data to follow the trend for more than four years to assess sustainability. That all three measures of bias decline after investigation suggests developed measures may correspond with other bias assessments.

[Figure 9. DOJ investigations over time.]

The slight drop during the investigation year suggests agencies might behave slightly better when DOJ is on site. However, the slight up-tick following investigation start is interesting and possibly indicates one of several processes. First, DOJ investigations take significant time, and it is not unusual for it to take more than a year before a court order or consent decree, creating a period of time after investigation but before structural changes are made to address bias. That the up-tick is sometimes slightly higher than original levels may indicate some backlash following investigation. However, we would again caution against over-interpretation given the sample size.

CONCLUSION

Racial bias in police arresting behavior is not only problematic for people of color, it is also a threat to the legitimacy of law enforcement as an institution. Because of the key role of police in the law enforcement pipeline, bias in arresting behavior has the potential to infiltrate the entire institution of justice. To date, social science has failed to develop a good measure of racial bias, inhibiting research on the causes of racial bias and limiting the ability of policy makers to effect change.

This paper develops three measures of racial disparity and bias in Black, Asian, and AIAN arrests, for 14,067 local, county and tribal agencies, between 1999 and 2015.

The first measure we develop is a simple logged risk ratio that takes the demographic composition of the arrestable population. The measure provides an easily interpreted measure of relative arrests. The risk of being arrested for all three populations increases during the time period. Arrest rates are a combination of two processes: (1) police policies and procedures that result in arrest; and (2) the actual rate at which crimes are committed. Since the risk ratio does not take into account racial differences in the underlying rate of commission of crime, it does not provide a measure of agency behavior. Therefore, it is best used when researchers have no need to distinguish between crime rates and policing behavior.

The second measure developed is a residual measure of arrest after controlling for explanatory variables which correlate to the commission of crime. The goal is to remove the portion of variance in arrest rates resulting from the actual crime rate, leaving the variance attributable to police action unexplained.

Across both adult and juvenile arrests, the residual for Black arrests is 0.020, for Asian arrests is 0.004, and for AIAN arrests is 0.007. However, there are a large number of agencies who arrest zero individuals of a certain race. If we exclude these agencies, numbers increase to 1.05, 2.19, and 3.40, respectively.

The third measure developed is a propensity-controlled residual measure that is more directly comparable across time and place than the risk ratio. However, because the measure relies on demographic and socio-economic variables that correlate with crime rates, it is not possible to then study how these same variables effect racial bias in arrests.

Also, the residual is a valid measure of racial disparity only to the extent it nets out the actual commission of crime. Since this is unobservable, the measure is sensitive to omitted variable bias. To help address this problem, the third measure standardizes the residual by the agency's jurisdiction (county, local, and tribal), type of offense, state, and year. The standardized measure is more comparable over time, as it helps account for the jurisdictional differences in the definition of crimes.

To help assess the validity of the measures, we utilize records of law enforcement misconduct investigations conducted by the Department of Justice, the gold standard in determining racial bias.

While the sample size is very small, all three measures exhibit higher bias in Black Arrests in instances where the DOJ found evidence of bias against African-Americans, compared to instances where bias allegations were made but not substantiated. Additionally, all three measures indicate bias declines after DOJ investigations.

All three measures increase between 1999 and 2015, a time when crime was generally decreasing. It remains an open question as to what caused this increase. We hope that the measures developed here might provide researchers and policy makers with tools needed to address racial bias and create police reform.

REFERENCES

Allport, Gordon W. 1954 (1979). *The Nature of Prejudice*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Andresen, Martin A. and Nicolas Malleson. 2013. "Crime Seasonality and Its Variations across Space." *Applied Geography* 43:25-35. doi: https://doi.org/10.1016/j.apgeog.2013.06.007.

Bachman, Ronet. 1998. "The Factors Related to Rape Reporting Behavior and Arrest: New Evidence from the National Crime Victimization Survey." *Criminal Justice and Behavior* 25(1):8-29. doi: 10.1177/0093854898025001002.

Blau, Judith R. and Peter M. Blau. 1982. "The Cost of Inequality: Metropolitan Structure and Violent Crime." *American Sociological Review* 47(1):114-29. doi: 10.2307/2095046.

Bollen, Kenneth A., Paul P. Biemer, Alan F. Karr, Stephen Tueller and Marcus E. Berzofsky. 2016. "Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis." *Annual Review of Statistics and its Application* 3(1):375-92. doi: 10.1146/annurev-statistics-011516-012958.

Bromwich Group LLC. 2016. "The Durability of Police Reform: The Metropolitan Police Department and Use of Force: 2008-2015." Vol.

Burch, Andrea M. 2007. "Sheriffs' Offices, 2007 - Statistical Tables." Vol. *Bureau of Justice Statistics: Statistical Tables.* Washington, D.C.

Chanin, Joshua M. 2014. "Examining the Sustainability of Pattern or Practice Police Misconduct Reform." *Police Quarterly* 18(2):163-92. doi: 10.1177/1098611114561305.

Davis, Robert C. , Nicole J. Henderson and Christopher W. Ortiz. 2005. "Can Federal Intervention Bring Lasting Improvement in Local Policing? The Pittsburgh Consent Decree." Vol.

de Lint, Willem. 2003. "Keeping Open Windows. Police as Access Brokers." *The British Journal of Criminology* 43(2):379-97. doi: 10.1093/bjc/43.2.379 %J The British Journal of Criminology.

Eschbach, Karl, Khalil Supple and C. Matthew Snipp. 1998. "Changes in Racial Identification and the Educational Attainment of American Indians, 1970-1990." *Demography* 35(1):35-43. doi: 10.2307/3004025.

Federal Bureau of Investigation, Uniform Crime Reporting Program. 2018, "Data Quality Guidelines". Retrieved July 1, 2018 (https://ucr.fbi.gov/data-quality-guidelines-new).

Fleiss, J.L. 1993. "Review Papers: The Statistical Basis of Meta-Analysis." *Statistical methods in medical research* 2(2):121-45.

Goff, Phillip Atiba and Kimberly Barsamian Kahn. 2012. "Racial Bias in Policing: Why We Know Less Than We Should." *Social Issues and Policy Review* 6(1):177-210. doi: 10.1111/j.1751-2409.2011.01039.x.

Harer, Miles D. and Darrell Steffensmeier. 1992. "The Differing Effects of Economic Inequality on Black and White Rates of Violence." *Social Forces* 70(4):1035-54. doi: 10.1093/sf/70.4.1035 %J Social Forces.

Harmon, Rachel. 2013. "Why Do We (Still) Lack Data on Policing." *Marquette Law Review* 96:1119-46.

Hartman, Jennifer L. and Joanne Belknap. 2003. "Beyond the Gatekeepers: Court Professionals' Self-Reported Attitudes About and Experiences with Misdemeanor Domestic Violence Cases." *Criminal Justice and Behavior* 30(3):349-73. doi: 10.1177/0093854803030003005.

Hashima, Patricia Y. and David Finkelhor. 1999. "Violent Victimization of Youth Versus Adults in the National Crime Victimization Survey." *Journal of Interpersonal Violence* 14(8):799-820. doi: 10.1177/088626099014008002.

Hindelang, Michael J. 1981. "Variations in Sex-Race-Age-Specific Incidence Rates of Offending." *American Sociological Review* 46(4):461-74. doi: 10.2307/2095265.

Justice, U.S. Department of and Federal Bureau of Investigation. 2010. *Crime in the United States, 2010: Caution against Ranking*Congress:3. Retrieved February 15, 2019 (https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/cautionagainstranking.pdf).

Justice, U.S. Department of. 2015. *Investigation of the Ferguson Police Department*Congress.

Krivo, Lauren J. and Ruth D. Peterson. 1996. "Extremely Disadvantaged Neighborhoods and Urban Crime." *Social Forces* 75(2):619-48. doi: 10.1093/sf/75.2.619 %J Social Forces.

Krosch, Amy R. and David M. Amodio. 2014. "Economic Scarcity Alters the Perception of Race." 111(25):9079-84. doi: 10.1073/pnas.1404448111 %J Proceedings of the National Academy of Sciences.

LaFree, Gary, Kriss A. Drass and Patrick O'Day. 1992. "Race and Crime in Postwar America: Determinants of African-American and White Rates, 1957-1988." *Criminology* 30(2):157-88. doi: 10.1111/j.1745-9125.1992.tb01101.x.

LaFree, Gary and Kriss A. Drass. 1996. "The Effect of Changes in Intraracial Income Inequality and Educational Attainment on Changes in Arrest Rates for African Americans and Whites, 1957 to 1990." *American Sociological Review* 61(4):614-34. doi: 10.2307/2096396.

LaFree, Gary, Eric P. Baumer and Robert O'Brien. 2010. "Still Separate and Unequal? A City-Level Analysis of the Black-White Gap in Homicide Arrests since 1960." *American Sociological Review* 75(1):75-100. doi: 10.1177/0003122409357045.

Land, Kenneth C., Patricia L. McCall and Lawrence E. Cohen. 1990. "Structural Covariates of Homicide Rates: Are There Any Invariances across Time and Social Space?". *American Journal of Sociology* 95(4):922-63. doi: 10.1086/229381.

Lapham, Sandra C., Betty J. Skipper, Iyiin Chang, Kerry Barton and Roderick Kennedy. 1998. "Factors Related to Miles Driven between Drinking and Arrest Locations among Convicted Drunk Drivers." *Accident Analysis & Prevention* 30(2):201-06. doi: https://doi.org/10.1016/S0001-4575(97)00084-5.

Leiber, Michael J., Mahesh K. Nalla and Margaret Farnworth. 1998. "Explaining Juveniles' Attitudes toward the Police." *Justice Quarterly* 15(1):151-74. doi: 10.1080/07418829800093671.

Liska, Allen E. and Mitchell B. Chamlin. 1984. "Social Structure and Crime Control among Macrosocial Units." *American Journal of Sociology* 90(2):383-95.

Little, Roderick J. and Sonya Vartivarian. 2003. "On Weighting the Rates in Non-Response Weights." 22(9):1589-99. doi: 10.1002/sim.1513.

Lott, John R. and John Whitley. 2003. "Measurement Error in County-Level Ucr Data." *Journal of Quantitative Criminology* 19(2):185-98. doi: 10.1023/a:1023054204615.

Maltz, Michael D. 1999. *Bridging Gaps in Police Crime Data*, Edited by U. S. D. o. Justice and O. o. J. Programs: DIANE publishing.

Maltz, Michael D. and Joseph %J Journal of Quantitative Criminology Targonski. 2002. "A Note on the Use of County-Level Ucr Data." *Journal of Quantitative Criminology* 18(3):297-318. doi: 10.1023/a:1016060020848.

McDowall, David, Colin Loftin and Matthew Pate. 2012. "Seasonal Cycles in Crime, and Their Variability." *Journal of Quantitative Criminology* 28(3):389-410. doi: 10.1007/s10940-011-9145-7.

Messner, Steven F. and Reid M. Golden. 1992. "Racial Inequality and Racially Disaggregated Homicide Rates: An Assessment of Alternative Theoretical Explanations." *Criminology* 30(3):421-48. doi: 10.1111/j.1745-9125.1992.tb01111.x.

Miethe, Terance D., Michael Hughes and David McDowall. 1991. "Social Change and Crime Rates: An Evaluation of Alternative Theoretical Approaches." *Social Forces* 70(1):165-85. doi: 10.2307/2580067.

Neapolitan, Jerome L. 1994. "Cross-National Variation in Homicides: The Case of Latin America." *International Criminal Justice Review* 4(1):4-22. doi: 10.1177/105756779400400102.

Nivette, Amy E. 2011. "Cross-National Predictors of Crime: A Meta-Analysis." *Homicide Studies* 15(2):103-31. doi: 10.1177/1088767911406397.

O'Brien, Robert M. 1996. "Police Productivity and Crime Rates, 1973-1992." *Criminology* 34(2):183-207. doi: 10.1111/j.1745-9125.1996.tb01202.x.

O'Brien, Robert M. 2003. "Ucr Violent Crime Rates, 1958–2000: Recorded and Offender-Generated Trends." *Social Science Research* 32(3):499-518. doi: https://doi.org/10.1016/S0049-089X(03)00020-6.

Osgood, D. Wayne. 2000. "Poisson-Based Regression Analysis of Aggregate Crime Rates." *Journal of Quantitative Criminology* 16(1):21-43. doi: 10.1023/A:1007521427059.

Ousey, Graham C. 1999. "Homicide, Structural Factors, and the Racial Invariance Assumption." *Criminology* 37(2):405-26. doi: 10.1111/j.1745-9125.1999.tb00491.x.

Parker, Karen F. and Patricia L. McCall. 1999. "Structural Conditions and Racial Homicide Patterns: A Look at Multiple Disadvantages in Urban Areas." *Criminology* 37(3):447-78. doi: 10.1111/j.1745-9125.1999.tb00493.x.

Peterson, Ruth D. and Lauren J. Krivo. 1993. "Racial Segregation and Black Urban Homicide." *Social Forces* 71(4):1001-26. doi: 10.1093/sf/71.4.1001 %J Social Forces.

Pfeffermann, Danny and Michail Sverchkov. 1999. "Parametric and Semi-Parametric Estimation of Regression Models Fitted to Survey Data." *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 61(1):166-86.

Phillips, Julie A. 1997. "Variation in African-American Homicide Rates: An Assessment of Potential Explanations." *Criminology* 35(4):527-60. doi: 10.1111/j.1745-9125.1997.tb01229.x.

Phillips, Julie A. 2002. "White, Black, and Latino Homicide Rates: Why the Difference?". *Social Problems* 49(3):349-73. doi: 10.1525/sp.2002.49.3.349.

Reaves, Brian A. 2015. "Local Police Departments, 2013: Personnel, Policies, and Practices." Vol. *Bulletin.* Washington, D.C.

Ross, Darrell L. and Patricia A. Parke. 2009. "Policing by Consent Decree: An Analysis of 42 U.S.C. § 14141 and the New Model for Police Accountability." *Police Practice and Research* 10(3):199-208. doi: 10.1080/15614260802381109.

Rushin, Stephen. 2014. "Structural Reform Litigation in American Police Departments." *Minnesota Law Review* 99:1343-422.

Sampson, Robert J. 1985. "Neighborhood and Crime: The Structural Determinants of Personal Victimization." *Journal of Research in Crime and Delinquency* 22(1):7-40. doi: 10.1177/0022427885022001002.

Sampson, Robert J. 1987. "Urban Black Violence: The Effect of Male Joblessness and Family Disruption." 93(2):348-82. doi: 10.1086/228748.

Sampson, Robert J. and William Julius Wilson. 1995. "Race, Crime, and Urban Inequality." Pp. 37-5 in *Crime and Inequality*, edited by J. Hagan and R. D. Peterson. Stanford, CA: Stanford University Press.

Shihadeh, Edward S. and Darrell J. Steffensmeier. 1994. "Economic Inequality, Family Disruption, and Urban Black Violence: Cities as Units of Stratification and Social Control." *Social Forces* 73(2):729-51. doi: 10.1093/sf/73.2.729 %J Social Forces.

Shihadeh, Edward S. and Nicole Flynn. 1996. "Segregation and Crime: The Effect of Black Social Isolation on the Rates of Black Urban Violence." *Social Forces* 74(4):1325-52. doi: 10.1093/sf/74.4.1325 %J Social Forces.

Shihadeh, Edward S. and Graham C. Ousey. 1996. "Metropolitan Expansion and Black Social Dislocation: The Link between Suburbanization and Center-City Crime." *Social Forces* 75(2):649-66. doi: 10.1093/sf/75.2.649 %J Social Forces.

Shihadeh, Edward S. and Graham C. Ousey. 1998. "Industrial Restructuring and Violence: The Link between Entry-Level Jobs, Economic Deprivation, and Black and White Homicide." *Social Forces* 77(1):185-206. doi: 10.1093/sf/77.1.185 %J Social Forces.

Skogan, Wesley G. 1982. "Methodological Issues in the Measurement of Crime." Pp. 203-08 in *The Victim in International Perspective* edited by H. J. Schneider. Berlin: Walter de Gruyter.

Skogan, Wesley G. and Kathleen Frydl, eds. 2004. *Fairness and Effectiveness in Policing: The Evidence*, Edited by C. o. L. a. J. National Research Council of the National Academies. Washington, D.C.: The National Academies Press.

Skogan, Wesley G. 2008. "Why Reforms Fail." *Policing and Society* 18(1):23-34. doi: 10.1080/10439460701718534.

Steffensmeier, Darrell and Haynie Dana. 2000. "Gender, Structural Disadvantage, and Urban Crime: Do Macrosocial Variables Also Explain Female Offending Rates?". *Criminology* 38(2):403-38. doi: 10.1111/j.1745-9125.2000.tb00895.x.

Stewart, Eric A., Eric P. Baumer, Rod K. Brunson and Ronald L. Simons. 2009. "Neighborhood Racial Context and Perceptions of Police-Based Racial Discrimination among Black Youth." *Criminology* 47(3):847-87. doi: 10.1111/j.1745-9125.2009.00159.x.

Stone, Christopher, Todd Foglesong and Christine M. Cole. 2009. "Policing Los Angeles under a Consent Decree: The Dynamics of Change at the Lapd." Vol. Cambridge, MA: Harvard Kennedy School.

Teplin, Linda A., Gary M. McClelland, Karen M. Abram and Dana A. Weiner. 2005. "Crime Victimization in Adults with Severe Mental Illness: Comparison with the National Crime Victimization Survey." *Archives of General Psychiatry* 62(8):911-21. doi: 10.1001/archpsyc.62.8.911 %J Archives of General Psychiatry.

Tyler, Tom R. and Yuen J. Huo. 2002. *Trust in the Law : Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Foundation.

U.S. Department of Justice and Federal Bureau of Investigation. 2004. "Uniform Crime Reporting Handbook." Vol. Clarksberg, WV.

U.S. Department of Justice, Civil Rights Division. 2017. "The Civil Rights Division's Pattern and Practice Police Reform Work: 1994-Present." Vol. January.

Walker, Samuel. 2007. "Police Accountability: Current Issues and Research Needs." Paper presented at the National Institute of Justice (NIJ) Policing Research Workshop: Planning for the Future, November 28-29, 2006, Washington, D.C.

Walker, Samuel. 2012. "Institutionalizing Police Accountability Reforms: The Problem of Making Police Reforms Endure." *Saint Louis University Public Law Review* 32(1):57-92.

Weitzer, Ronald and Steven A. Tuch. 2004. "Race and Perceptions of Police Misconduct." *Social Problems* 51(3):305-25. doi: 10.1525/sp.2004.51.3.305 %J Social Problems.

Weitzer, Ronald and Steven A. Tuch. 2005. "Determinants of Public Satisfaction with the Police." 8(3):279-97. doi: 10.1177/1098611104271106.
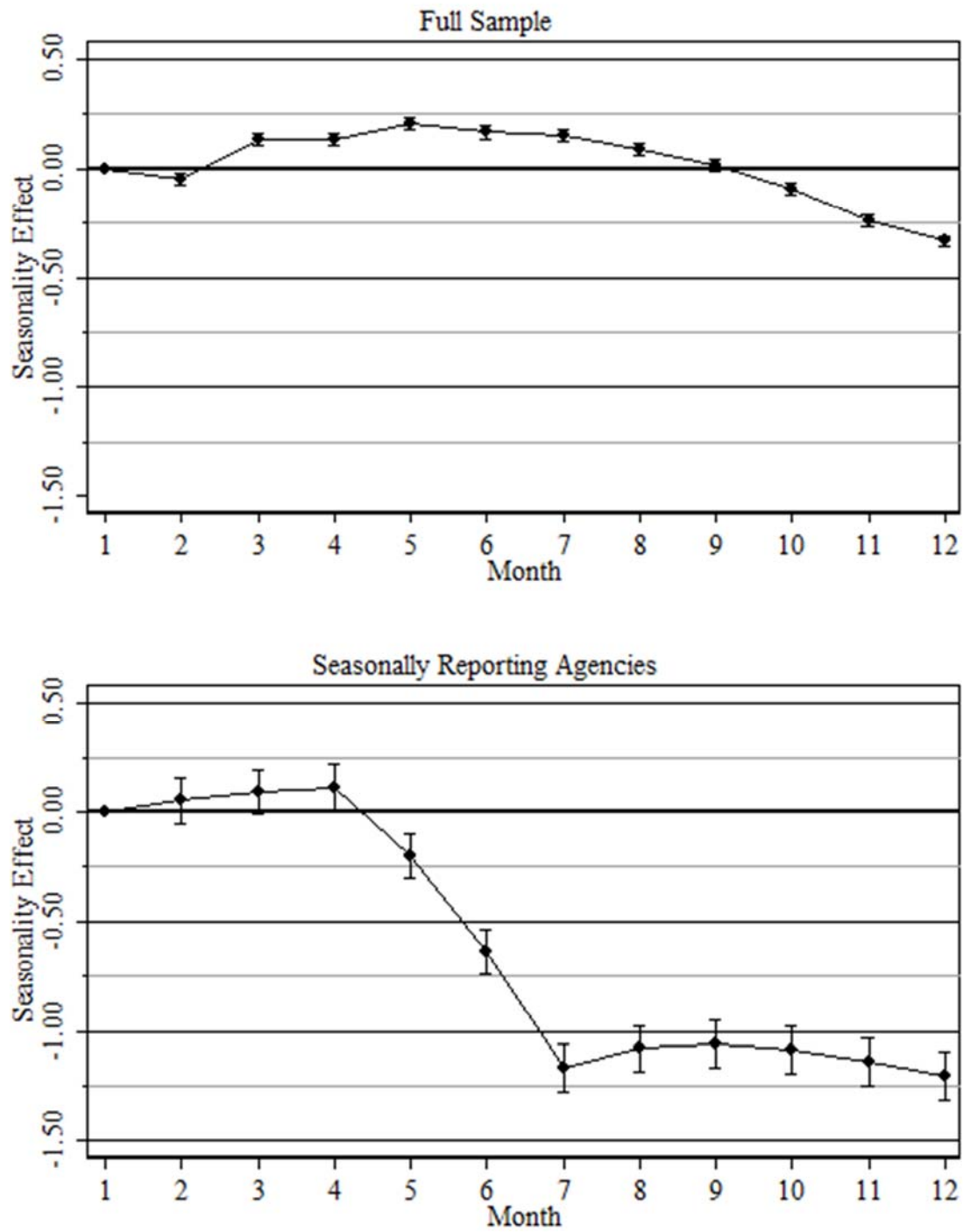
FIG. 1 – Average seasonality effect for full agency sample and for
seasonally reporting agencies, over full period.
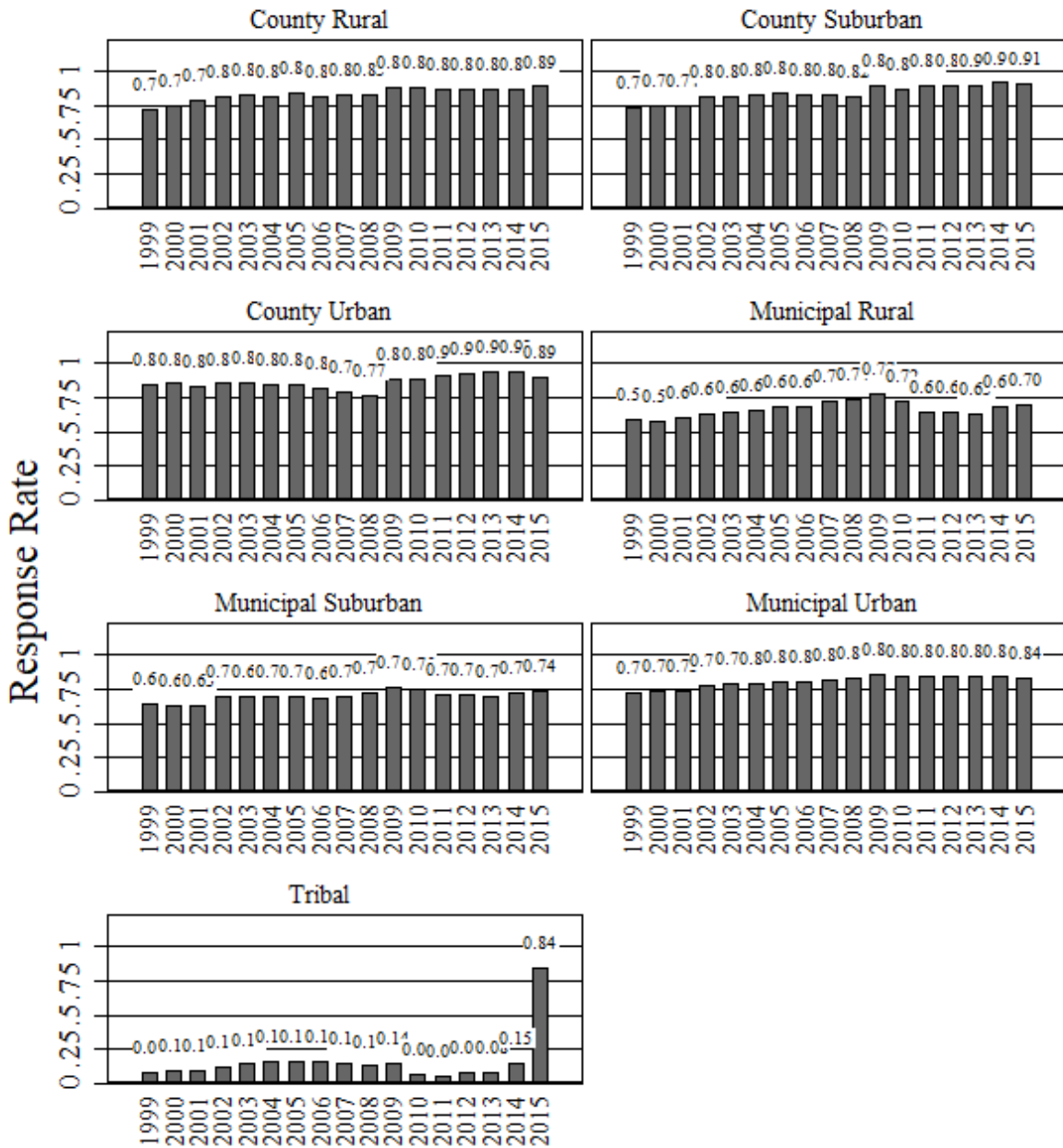
NOTES. – Bars denote 95% confidence interval.

FIG. 2 – Agency response rates, by weighting strata, 1999-2015.

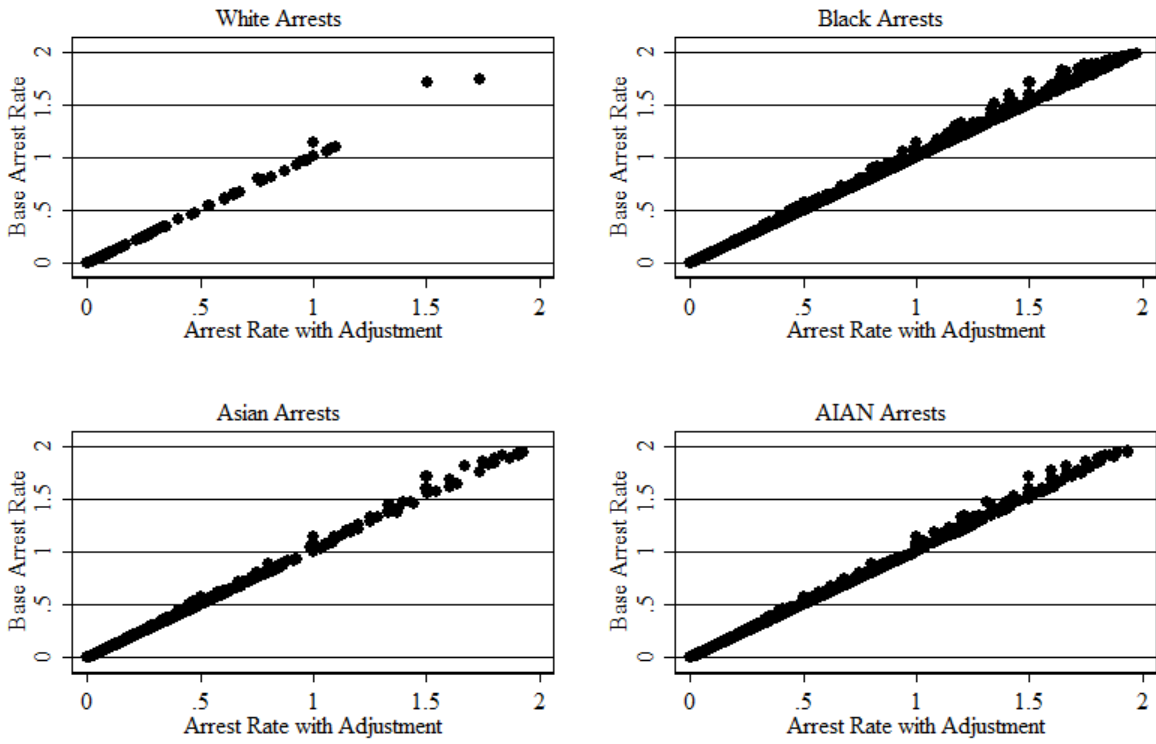NOTES. – Agencies reporting fewer than three months are counted as non-reporting.

FIG. 3 – Base and adjusted arrest rates.

NOTES. – Base rates are calculated as Total Arrests / Population. Adjusted rates are calculated as (Total Arrests + Population Proportion) / (Population +1)
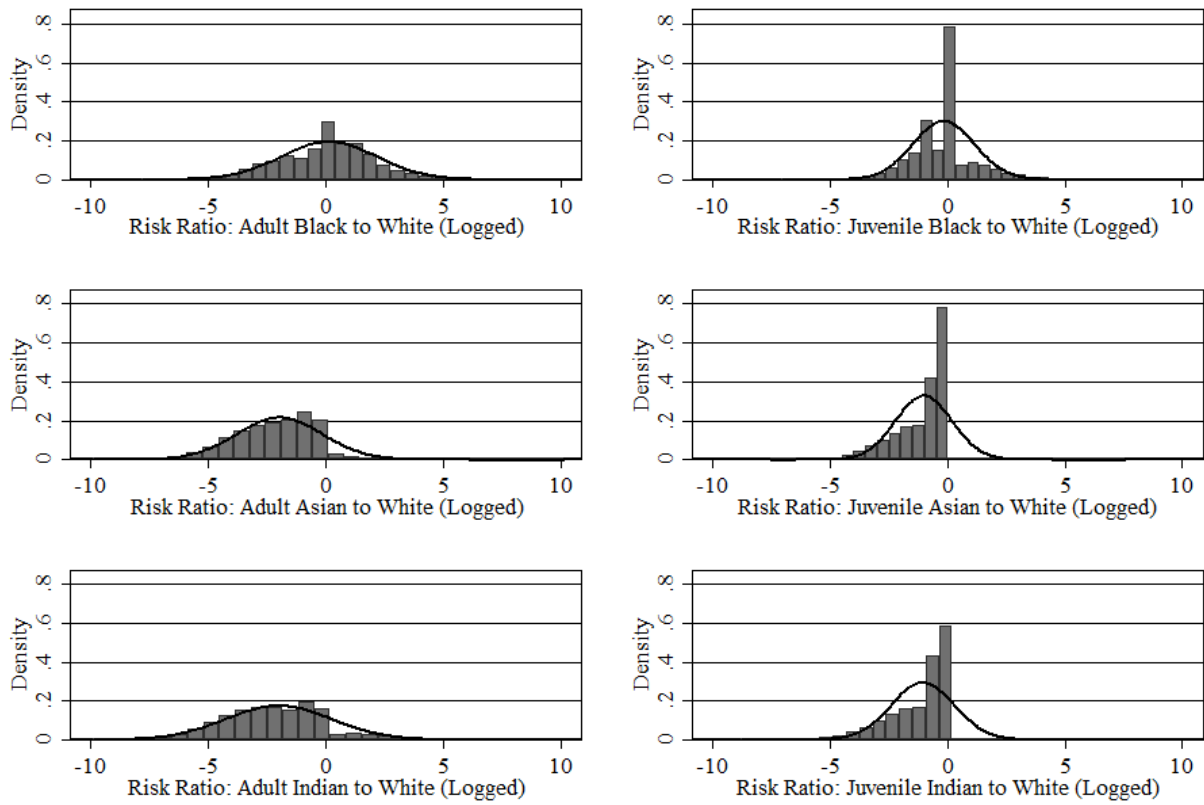
Fɪɢ. 4 – Distribution of risk ratio for adult and juvenile arrests, by race.
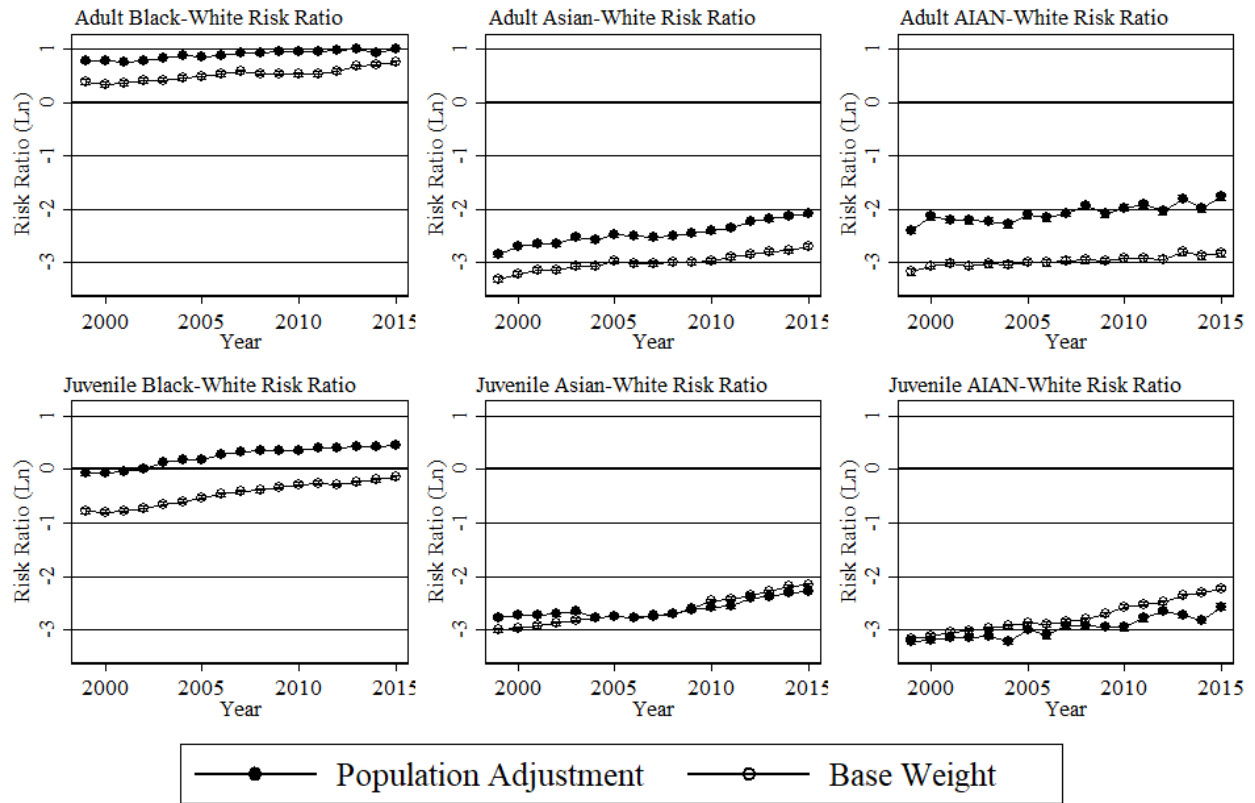
Nᴏᴛᴇs. – Line denotes normal-density function.

FIG. 5 – Weighted risk ratio for adult and juvenile arrests, by race.

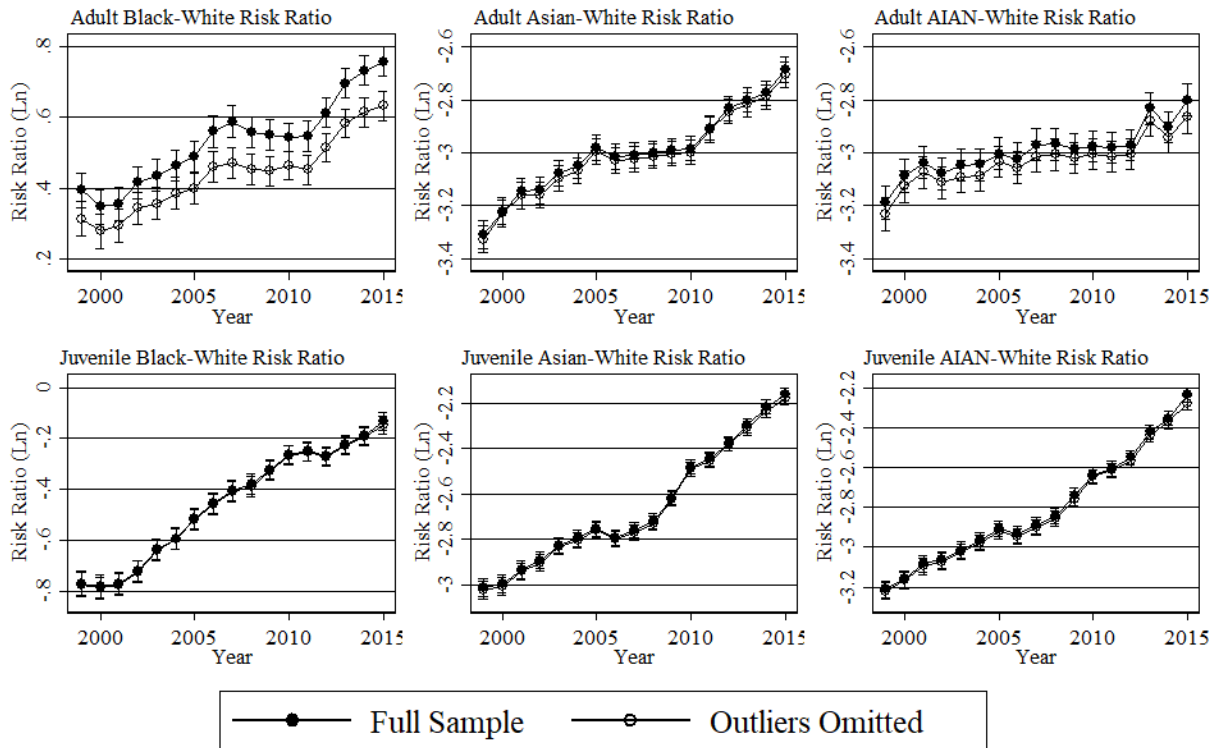NOTES. – Bars denote 95% confidence interval.

FIG. 6 – Mean risk ratio for adult and juvenile arrests, estimated with full sample or with outliers omitted, by race.
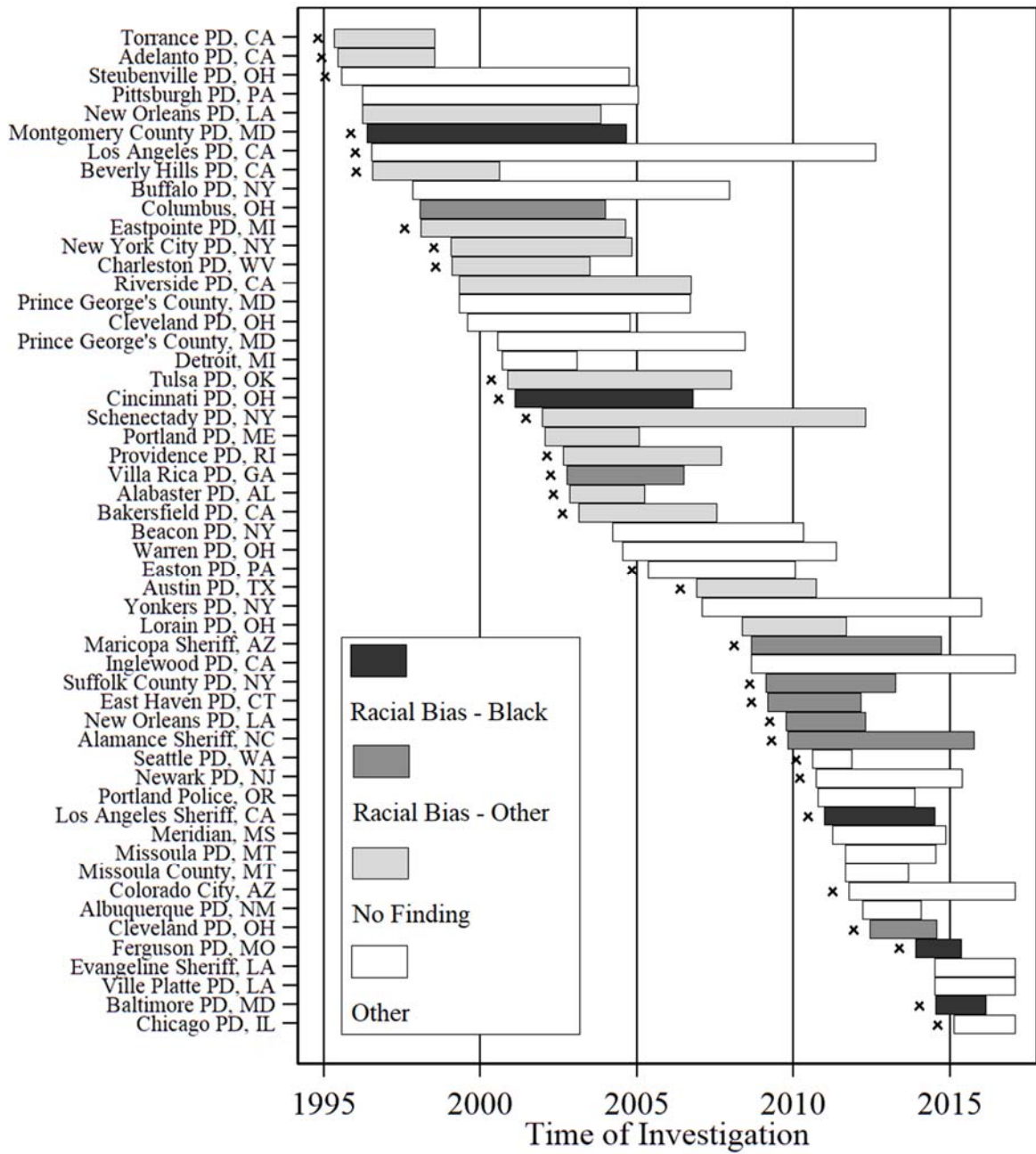
NOTES. – Bars denote 95% confidence interval.

FIG. 7 – DOJ investigation start and end dates, by finding.

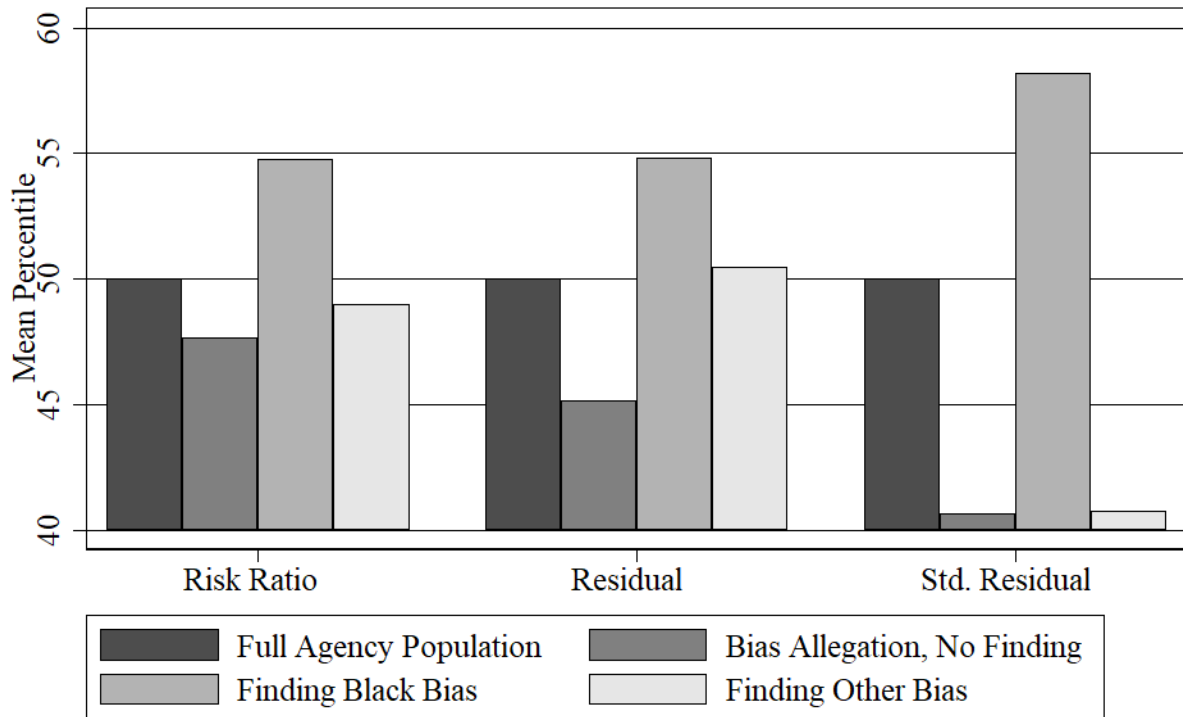NOTES. – x Denotes investigations with racial bias allegations.

FIG. 8 – Mean percentile for measures of Black-White racial disparity, by DOJ investigation finding.

NOTES. – 'Full Agency Population' category includes agencies with no investigations and investigations with no racial bias allegation. 'Bias Allegation, No Finding' includes 10 agencies. Investigations with finding of bias against Blacks includes five (5) agencies. Investigations with findings of other racial bias includes seven (7) agencies.
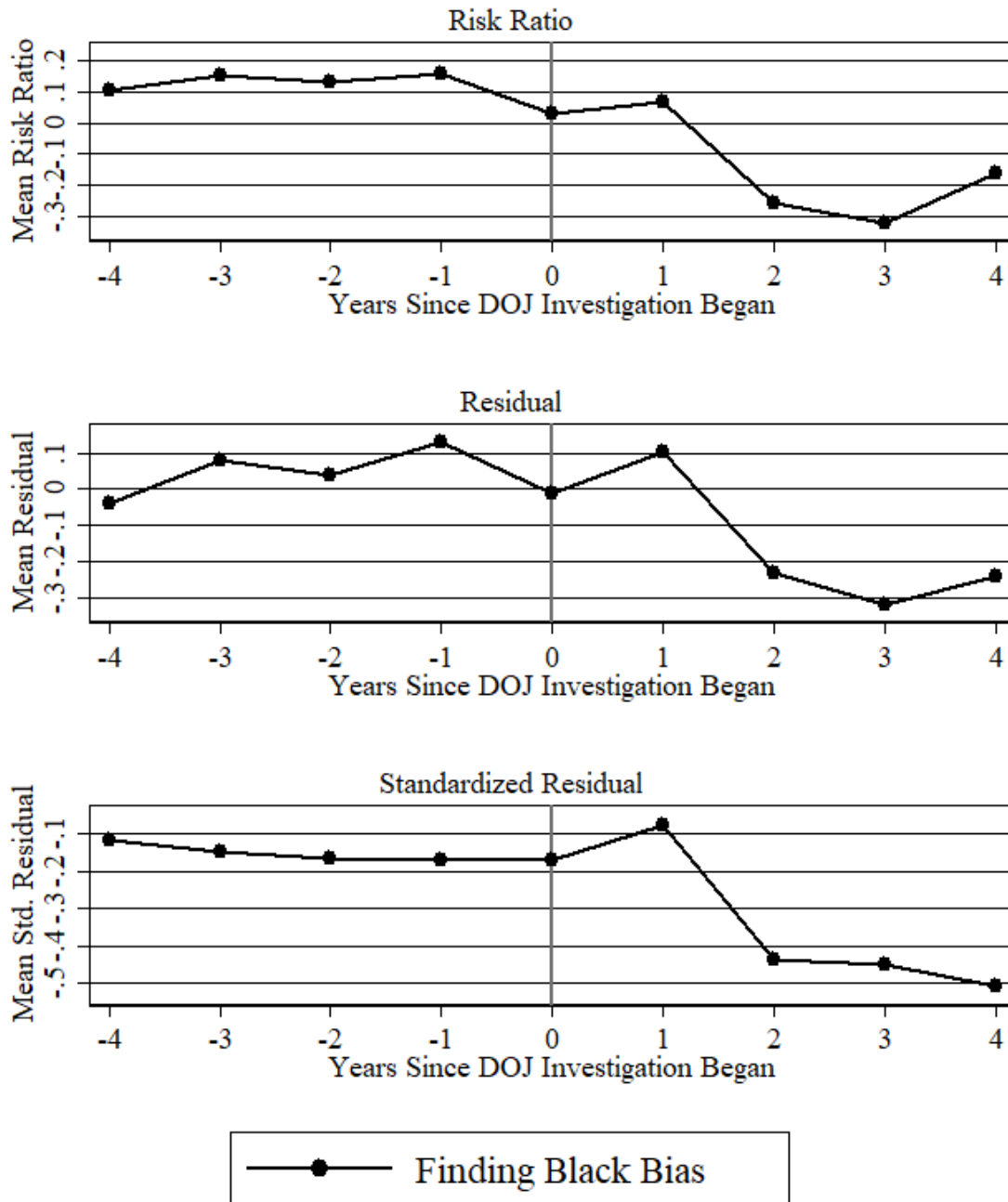
FIG. 9 – Difference-in-Difference changes in Black-White racial disparity, by DOJ investigation finding.

NOTES. – DOJ investigation begins in year 0. Allegation with no finding includes 10 agencies. Investigations with finding of bias against Blacks includes five (5) agencies.

TABLE 10 – Variables empirically demonstrated to correlate with crime rates.

Variable	Source
Income Inequality and Insecurity	
Inequality	(Blau and Blau 1982, Harer and Steffensmeier 1992, Ousey 1999, Shihadeh and Steffensmeier 1994, Shihadeh and Flynn 1996)
Black-White Income Ratio	(Harer and Steffensmeier 1992, Messner and Golden 1992, Parker and McCall 1999, Shihadeh and Steffensmeier 1994, Shihadeh and Flynn 1996)
Poverty Rate	(Krivo and Peterson 1996, Ousey 1999, Shihadeh and Ousey 1998)
Family Income	(LaFree and Drass 1996)
Male Income	(LaFree and Drass 1996)
Earnings	(Phillips 1997)
Deprivation Index	(Land, McCall and Cohen 1990, Ousey 1999)
Family Disruption	
Female headed households	(LaFree, Drass and O'Day 1992, Ousey 1999, Phillips 1997, Sampson 1985, Sampson 1987, Shihadeh and Steffensmeier 1994)
Children not in two parent households	(Parker and McCall 1999)
Divorce rate	(Land, McCall and Cohen 1990, Phillips 1997)
Male divorce rate	(Parker and McCall 1999)
Female fertility rate, age 15-19	(Phillips 1997)
Geographic mobility	(Phillips 1997, Sampson 1985)
Residential Segregation	
Index of Dissimilarity	(Messner and Golden 1992, Parker and McCall 1999, Peterson and Krivo 1993)
Geographical Spread	(Shihadeh and Flynn 1996)
Suburban Residency Rate	(Shihadeh and Flynn 1996, Shihadeh and Ousey 1996)

Educational Attainment	(LaFree, Drass and O'Day 1992, LaFree and Drass 1996, Messner and Golden 1992, Parker and McCall 1999, Phillips 1997)
Male educational attainment	(LaFree, Drass and O'Day 1992)
Labor Market Attachment	
Joblessness	(Parker and McCall 1999, Phillips 1997, Sampson 1987)
Male Joblessness	(Krivo and Peterson 1996, LaFree, Drass and O'Day 1992)
Occupational Composition	(Krivo and Peterson 1996, Phillips 1997) (Shihadeh and Ousey 1998)
Labor Participation	(LaFree and Drass 1996, Phillips 1997)
Unemployment rate	(Messner and Golden 1992, Parker and McCall 1999, Phillips 1997, Shihadeh and Ousey 1998)
Social Organization	
Density of social networks	(Sampson 1985, Sampson and Wilson 1995)
Intergenerational links	(Sampson and Wilson 1995)
Street-Corner Peer Groups	(Sampson and Wilson 1995)
Organizational participation	(Sampson and Wilson 1995)
Probability of Black-White Interaction	(Shihadeh and Flynn 1996)