



ARTICLE



<https://doi.org/10.1057/s41599-023-01879-5>

OPEN

Federal criminal sentencing: race-based disparate impact and differential treatment in judicial districts

Chad M. Topaz ^{1,2,3✉}, Shaoyang Ning¹, Maria-Veronica Ciocanel^{2,4} & Shawn Bushway^{5,6}

Race-based inequity in federal criminal sentencing is widely acknowledged, and yet our understanding of it is far from complete. Inequity may arise from several sources, including direct bias of courtroom actors and structural bias that produces racially disparate impacts. Irrespective of these sources, inequity may also originate from different loci within the federal system. We bring together the questions of the sources and loci of inequity. The purpose of our study is to quantify race-based disparate impact and differential treatment at the national level and at the level of individual federal judicial districts. We analyze over one-half million sentencing records publicly available from the United States Sentencing Commission database, spanning the years 2006 to 2020. At the system-wide level, Black and Hispanic defendants receive average sentences that are approximately 19 months longer and 5 months longer, respectively. Demographic factors and sentencing guideline elements account for nearly 17 of the 19 months for Black defendants and all five of the months for Hispanic defendants, demonstrating the disparate impact of the system at the national level. At the individual district level, even after controlling for each district's unique demographics and implementation of sentencing factors, 14 districts show significant differences for minoritized defendants as compared to white ones. These unexplained differences are evidence of possible differential treatment by judges, prosecutors, and defense attorneys.

¹Williams College, Williamstown, Massachusetts, USA. ²Institute for the Quantitative Study of Inclusion, Diversity, and Equity, Williamstown, Massachusetts, USA. ³University of Colorado, Boulder, USA. ⁴Duke University, Durham, North Carolina, USA. ⁵The RAND Corporation, Santa Monica, California, USA. ⁶State University of New York, Albany, New York, USA. ✉email: chad@qsideinstitute.org

Introduction

In the United States, the federal system of criminal sentencing was created to ensure that federal courts provide similar sentences for similar criminal conduct by individuals with similar criminal histories. At the same time, according to the judicial branch itself, there exists substantial racial inequity in sentencing (Franklin, 2013). As compared to similarly situated white men, Hispanic and Black men receive sentences that are, respectively, 5% and 20% longer on average (United States Sentencing Commission, 2018). While the disparity between Black and white defendants is particularly well-established (Feldmeyer and Ulmer, 2011; Mustard, 2001; Rachlinski and Wistrich, 2017), there is compelling evidence that Native and Hispanic men—especially young men—also receive harsher sentences (Doerner and Demuth, 2010; Franklin, 2013).

How can such inequities exist alongside a system designed specifically to eliminate them? There are at least three reasons. First, the sentence recommended for an individual is determined primarily by their criminal history and the type of crime for which they have been convicted. This system can have a disparate impact on minoritized populations as compared to white ones if minoritized individuals have more substantial criminal histories or are likely to be convicted of more severe crimes. Second, some federal districts may apply sentencing guidelines in a manner that implicitly penalizes racially minoritized defendants as compared to white ones (Spohn, 2000; Yang, 2015). For instance, judges in a particular district may tend to use the upper end of presumptive sentencing ranges for certain offenses, adopting a stance such as “We’re tough on drugs in this district.” This type of structural differential impact could potentially amplify the first type. Finally, actors in the federal system may treat members of racial groups differently even after accounting for legal factors (Spohn, 2000). We are concerned with all three of the aforementioned sources of racial disparity. We will estimate the structural disparate impact and differential treatment by courtroom actors at the national level and, for the first time, within individual federal judicial districts.

How federal sentencing works. The Sentencing Reform Act of 1984 created the United States Sentencing Commission (USSC) and charged it with promoting equity in the justice system. To this end, USSC created its sentencing guidelines and put them into practice three years later (United States Sentencing Commission, 1987). The guidelines account for a defendant’s criminal history and the severity of their alleged crime, as well as numerous possible mitigating and aggravating factors. Based on these factors, each case is placed into a group, visualized as a cell in a sentencing grid, and each cell carries a specific range of time recommended for incarceration. One may view the guidelines as an algorithm that outputs a recommended sentencing range based on the characteristics of the defendant and on the alleged crime.

Initially, federal judges were required to follow these guidelines; that is, they were limited to imposing sentences that were within the recommended ranges. A pivotal change occurred after the Supreme Court case *United States v. Booker*, in which the Supreme Court held that mandatory guidelines violated the Sixth Amendment. As a result, the language making the guidelines mandatory was stripped as of early 2005, and the guidelines were thenceforth advisory. Legal scholarship has sought to assess the impact of *Booker* on racial disparities, largely with a focus on Black defendants. Some evidence suggests that Black-white disparities in sentencing increased after *Booker* (Yang, 2015), while other evidence indicates no such increase (Starr, 2013). Irrespective of *Booker*, some studies have found evidence

suggesting a decrease in race-based sentencing disparity over time (King and Light, 2019; Light, 2022).

Disparate impact and differential treatment. Though scholars debate whether *Booker* increased race-based sentencing disparities, there is little debate on whether the disparities existed both before and after the decision. Such disparities also show up in state sentencing (Abrams et al., 2012) and state bail decisions (Arnold et al., 2018). In thinking about sentencing differences due to defendant race, we distinguish between the differential impact of structural factors and differential treatment by actors in the courtroom workgroup, including judges, defense attorneys, and prosecutors (Bushway and Piehl, 2001).

Differential treatment by the actors in the courtroom workgroup has been a focus of research in this space, with much of the work in criminology guided by the focal concerns framework (Steffensmeier et al., 1998). This theory proposes that members of the courtroom workgroup have three focal concerns when determining appropriate sentences. The three focal concerns are the defendant’s blameworthiness, the extent to which a longer sentence would protect the community, and the practical constraints associated with a longer sentence. This framework says that members of the courtroom workgroup consciously or subconsciously rely on stereotypes to assess the three concerns, in turn reducing the cognitive demand of sentencing. For instance, some groups may be stereotyped as tough enough to handle prison, as being already absent from one’s dependents, and so forth. White defendants benefit from these stereotypes (Smith et al., 2014). Regarding blameworthiness, white individuals are underrepresented as perpetrators of crime in television news reports (Dixon and Linz, 2000), and this underrepresentation appears to provide white individuals an innocence premium in people’s minds (Dixon, 2006). Regarding danger to the community, white individuals are differentially perceived as safe and disinclined to commit crimes (Smith et al., 2014). Regarding practical constraints, qualitative evidence suggests that judges perceive white defendants and their families as being especially harmed by prison time (Kramer and Steffensmeier, 1993).

This racially disparate treatment is important but may account for a relatively small part of the overall racial disparity in the system (Baumer, 2013; Mitchell, 2005). Instead, the bulk of the racial disparity may be attributable to the structural properties of the criminal justice system (Bushway and Forst, 2013; Bushway and Piehl, 2011; Frase and Roberts, 2020). The guidelines system itself is an example of a structure that could create inequity since the recommended punishments were created based on past practices, which could well have been racially biased. Above and beyond the basic guidelines, other mitigating and aggravating factors can cause disparity. For example, the Anti-Drug Abuse Act of 1986 put in place mandatory minimum sentences for certain drug-related crimes. Twenty years after the passage of the law, Black defendants made up 15% of drug users in the U.S., 74% of individuals sentenced to prison for a drug offense, and over 80% of individuals sentenced for crack-related offenses (Vagins and McCurdy, 2006).

The disparity might also stem from variation among federal districts, each of which may choose to apply the sentencing guidelines in its own way. Documentation from USSC at the time they created the guidelines states that concerns about interdistrict variation were one of the primary motivations for the Sentencing Reform Act of 1984. The sentence given to a defendant results from the decisions of prosecutors, defense attorneys, and judges, and there is no reason to assume that the approaches of these actors are uniform across districts. Intra-city/inter-judge variation

in sentencing is wide and has generally grown post-*Booker* (Scott, 2010; Yang, 2014).

Unfortunately, the literature on intra-city (including inter-judge) and interdistrict differences is sparse. Some work in criminology has used multi-level models to demonstrate district-level variations in practices like the provision of downward departures (Johnson et al., 2008). USSC itself has released a report on interdistrict variation in sentencing (United States Sentencing Commission, 2020a). That said, the government's publicly released sentencing records do not attach judge names to individual sentences, thereby precluding judge-level analysis. The aforementioned judge-level studies (Scott, 2010; Yang, 2014) were possible only because they used proprietary data. However, the data usage agreement prohibits public access and disclosure of judge names.

Summary of results. In this paper, we estimate the structural disparate impact and differential treatment by courtroom actors at the national level and, for the first time, within individual federal judicial districts. We study over one-half million sentencing records from the USSC database, spanning the years 2006 to 2020. Black and Hispanic defendants receive average sentences that are approximately 19 months longer and 5 months longer, respectively. Personal demographics and sentencing guideline elements account for nearly 17 of the 19 months for Black defendants and all five of the months for Hispanic defendants, demonstrating the disparate impact of the system at the national level. To further investigate the source of these differences, we analyze individual districts. Assuming uniform treatment of sentencing variables across all districts, 22 districts display significantly longer conditional sentences for minoritized defendants. These disparities are evidence of either structural inequity and/or differential treatment within the district. However, even after controlling for each district's unique demographics and implementation of sentencing factors, 14 districts show significant differences for minoritized defendants as compared to white ones. These unexplained differences are evidence of possible differential treatment by judges, prosecutors, and defense attorneys.

Methodology

Data. We estimate race-based sentencing disparities from public data released by USSC (United States Sentencing Commission, 2021). To avoid potential confounding due to the *Booker* Supreme Court case in 2005, we utilize data from fiscal years 2006 through 2020, comprising 1,027,976 sentences before restriction. The codebook for these data (United States Sentencing Commission, 2020b) provides detailed explanations of all variables.

We immediately eliminate immigration cases using the variables that specify offense type (OFFGUIDE = 17, OFFTYPE2 = 27, or OFFTYP2B = 27, depending on the sentencing year) and any other cases involving noncitizen defendants (NEWCIT = 1). This is the same approach taken by, for example, Rehavi and Starr (2014), who state that "immigration cases and other cases involving noncitizens were excluded because the stakes in those cases center on deportation, making them not directly comparable to other crimes." There are 479,347 records that involve noncitizens and/or immigration cases and/or have missing citizenship data. After discarding these, 548,629 sentencing records remain. It is important to keep in mind that this exclusion of data means that the research questions we can address are necessarily limited in scope.

In our regression models, the outcome variable is the total prison sentence in months as captured by the variable SENTTOT in USSC data. There are 82,101 records (15.0%) with missing data for this variable. Among the records with missing data, the vast

majority, 75,338 records, have penalties involving probation, monetary fines, and/or home confinement. For these records, we convert the value of SENTTOT to zero months. After this imputation, there are still 6763 records with missing data for SENTTOT. These comprise 1.2% of the 548,629 records, and we eliminate them since we cannot determine the outcome in terms of prison sentence. At this stage, 541,866 records remain.

USSC codes life sentences as 470 months. There are 2295 such cases (0.4%) in the remaining data and we retain these. At the same time, 1570 records (0.3%) indicate sentences longer than 470 months. We discard this small amount of data in order to avoid a potentially nonmonotonic representation of prison sentences. At this stage, 540,296 records remain.

We will estimate the effects of defendant race and judicial district after controlling for various other explanatory variables. Our strategy is to focus primarily on variables directly related to sentencing, but also to include a small number of extra-legal variables, namely age, sex, and educational attainment. One could choose to exclude any and all extra-legal variables. Instead, we opt to include the few extra-legal variables commonly used as controls in the research literature and in the government's own studies. Starr (2013), Rehavi and Starr (2014), and United States Sentencing Commission (2018) are merely a few examples.

We eliminate any remaining records that are missing data for variables key to our analysis. These are the defendant characteristics race (NEWRACE), age (AGE), sex (MONSEX), and educational attainment (NEWEDUC), as well as the following sentencing variables: offense level after Chapter 2 adjustments (CHAP2), offense level after Chapter 3 adjustments (COADJLEV), final offense level (XFOLSOR), criminal history group (XCRHISSR), minimum of guideline sentencing range (XMINSOR), trumped guideline minimum (GLMIN), whether the defendant pleaded guilty (NEWCNVTN), and presence of a government departure based on the *Booker* reporting category (BOOKERCD through fiscal year 2017, and SENTRNGE afterward). There are 21,477 cases with missing data, comprising 4.0% of the data at the previous stage of elimination. We drop these from our analysis dataset, leaving 518,819 records.

Finally, we have only 98 records from the District of the Northern Mariana Islands, which is insufficient for estimating effects within that district. We eliminate those records from our study, leaving us with a final dataset of 518,721 records.

Treatment of variables. Except for defendant age (AGE) and prison sentence (SENTTOT), we treat all variables as categorical, regardless of coding that appears numerical in the raw data. This is especially important for offense-level variables (CHAP2, COADJLEV, XFOLSOR) and criminal history group (XCRHISSR). We make this choice because the meanings of these variables are not linear in their codings. For instance, a crime with offense level six is in no way expected to confer double the sentence of a crime with offense level three.

To further prepare our data for analysis, we add several derived variables. To track the application of mandatory minimum sentences, we create a flag (MANDMIN) that is true whenever the trumped guideline minimum (GLMIN) exceeds the guideline minimum (XMINSOR). To track government-sponsored downward departures, we create a flag (GOVDEPART) that is true whenever such a departure is coded in the relevant sentencing variables (BOOKERCD or SENTRNGE, depending on the year). Finally, as described in the "Introduction", a defendant's presumptive sentencing range is determined by the grid cell on which their alleged offense falls, which in turn is determined by the final, adjusted offense level (XFOLSOR) and by the criminal history group (XCRHISSR). For convenience, we create a

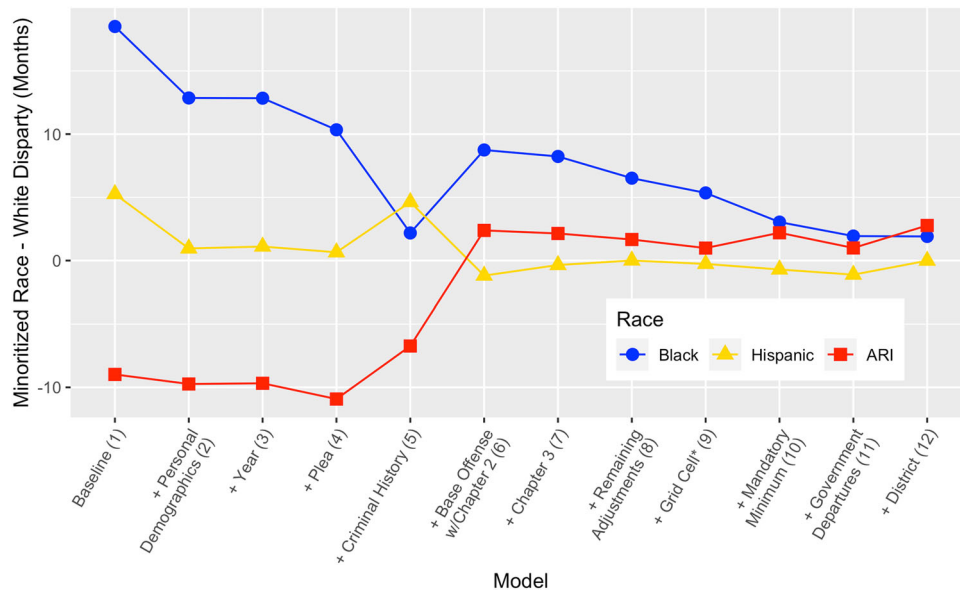


Fig. 1 Race-based sentencing disparity in the federal court system. Each point represents mean sentencing disparity (in months) as estimated from a regression model that conditions the outcome by adding an additional explanatory variable (given on the x-axis) to the model immediately to the left. The one exception is the “Grid Cell” model, marked with an asterisk (*) on the x-axis. This model combines the previous four explanatory variables into a single one representing the cell of the United States Sentencing Grid onto which each defendant is placed, as the grid cell sets the presumptive sentence. Disparities are shown as the average difference in sentences given to defendants who are Black (blue circles), Hispanic (yellow triangles), and another racial identity (ARI, red squares), each as compared to white defendants. The leftmost model, labeled “Baseline,” provides the average difference in sentences between racial groups irrespective of any and all other factors. The rightmost model demonstrates average disparities for minoritized defendants—especially Black and ARI ones—that remain unexplained. Table 1 reports the values of the estimates represented here along with *p*-values and confidence intervals. Table 2 provides a more detailed specification of the 12 regressions. The analysis is based on *N* = 518,721 sentencing records. Model diagnostics, including *F*-statistics, the associated degrees of freedom, the associated *p*-value, as well as raw and adjusted *r*² values appear in our permanent data repository (Topaz, 2023). Each of these *F*-statistic *p*-values is numerically indistinguishable from zero, and the adjusted *r*² values range from 0.02 for Model (1) up to 0.79 for Model (12).

Table 1 Race-based sentencing disparity in the federal court system.

Model (Model number)	Minoritized race-white disparity (months)		
	Black	Hispanic	ARI
Baseline (1)	18.5 ± 0.5****	5.3 ± 0.5****	-9.0 ± 0.9****
+ Personal demographics (2)	12.9 ± 0.5****	1.0 ± 0.5**	-9.7 ± 0.9****
+ Year (3)	12.8 ± 0.5****	1.1 ± 0.5***	-9.7 ± 0.9****
+ Plea (4)	10.3 ± 0.4****	0.7 ± 0.5	-10.9 ± 0.8****
+ Criminal history (5)	2.2 ± 0.4****	4.6 ± 0.5****	-6.7 ± 0.8****
+ Base offense w/Chapt. 2 (6)	8.7 ± 0.3****	-1.2 ± 0.3****	2.4 ± 0.6****
+ Chapter 3 (7)	8.2 ± 0.3****	-0.4 ± 0.3	2.1 ± 0.5****
+ Remaining adjustments (8)	6.5 ± 0.3****	0.0 ± 0.3	1.7 ± 0.5****
+ Grid cell (9)	5.4 ± 0.3****	-0.3 ± 0.3	1.0 ± 0.5****
+ Mandatory minimum (10)	3.0 ± 0.2****	-0.7 ± 0.3**	2.2 ± 0.5****
+ Government departures (11)	1.9 ± 0.2****	-1.1 ± 0.3****	1.0 ± 0.4****
+ District(12)	1.9 ± 0.2****	0.0 ± 0.3	2.8 ± 0.5****

Here we provide the numerical values underlying Fig. 1. Each row represents mean sentencing disparity (in months, along with a 95% confidence interval, obtained using robust standard errors) as estimated from a regression model that conditions the outcome by adding an additional explanatory variable to the model immediately above. The one exception is the “Grid cell” model. This model combines the previous four explanatory variables into a single one representing the cell of the United States Sentencing Grid onto which each defendant is placed, as grid cell sets the presumptive sentence. Table 2 provides a more detailed specification of each model. Disparities are given as the average difference in sentences given to defendants who are Black, Hispanic, and another racial identity (ARI), each as compared to white defendants. The top model, labeled “Baseline,” provides the average difference in sentences between racial groups irrespective of any and all other factors. The bottom model demonstrates average disparities for minoritized defendants that remain unexplained. Asterisks denote the range for a Bonferroni-adjusted *p*-value: * (*p* < 0.05), ** (*p* < 0.01), *** (*p* < 0.001), **** (*p* < 0.0001). Estimates with no asterisk were not found to be statistically significant (at a level of 0.05). The analysis is based on *N* = 518,721 sentencing records. Model diagnostics, including *F*-statistics, the associated degrees of freedom, the associated *p*-value, as well as raw and adjusted *r*² values appear in our permanent data repository (Topaz, 2023). Each of these *F*-statistic *p*-values is numerically indistinguishable from zero, and the adjusted *r*² values range from 0.02 for Model (1) up to 0.79 for Model (12).

categorical variable (GRID) out of all possible combinations of those two variables.

Statistical modeling. The results of Fig. 1 and Table 1 are based on linear regression models conducted in the statistical

computing package R. Here we make the assumptions that the various factors impacting the sentence are linearly additive and interactions between factors are negligible except as specified. The response variable is prison sentence (SENTTOT) and the explanatory variables are defendant race (NEWRACE) along with all the variables described in the captions. Table 2 provides a more

Table 2 Specification of regression models.

	Regression model											
	1	2	3	4	5	6	7	8	9	10	11	12
NEWRACE	●	●	●	●	●	●	●	●	●	●	●	●
AGE		●	●	●	●	●	●	●	●	●	●	●
MONSEX		●	●	●	●	●	●	●	●	●	●	●
NEWEDUC		●	●	●	●	●	●	●	●	●	●	●
SENTYR			●	●	●	●	●	●	●	●	●	●
NEWCVNTN				●	●	●	●	●	●	●	●	●
XCRHISSR					●	●	●	●				
CHAP2						●						
COADJLEV							●					
XFOLSOR								●				
GRID (XCRHISSR:XFOLSOR)									●	●	●	●
MANDMIN										●	●	●
GOVDEPART											●	●
CIRCDIST												●

The results summarized in Fig. 1 and Table 1 come from these 12 linear regressions. For each model, the response variable is the length of the sentence, SENTTOT. The explanatory variables included in each model are those marked with a bullet in the table below. See “Methodology” for a discussion of variable meanings.

detailed specification of the variables involved in each model. For each regression, we measure racial disparity via the difference between each coefficient for minoritized race (Black, Hispanic, ARI) and the coefficient for white race. Motivated by a visual inspection of the residuals, we use heteroskedasticity-robust standard errors to associate *p*-values, which we tabulate along with the estimates. As we have three measurements of disparity for each of the 12 models, there are 48 estimates and we Bonferroni adjust *p*-values for the differences.

The results for Figs. 2–4 and Table 3 are based on a separate set of linear regression models. District Model I is similar to Model 12 of Fig. 1 and Tables 1 and 2, except we replace district and race (as separate terms) with the interaction of district and race (treated categorically), that is:

$$SENTTOT \sim AGE + MONSEX + NEWEDUC + SENTYR + NEWCVNTN + GRID + MANDMIN + GOVDEPART + NEWRACE : CIRCDIST$$

District Model II takes the same regression formula as Model 11 of Fig. 1 and Tables 1 and 2 but applies it to each district individually. By using this framework, we allow district-specific impacts from various factors included in the model. For both district models, we again use robust standard errors. As we have three measurements of disparity for each of the 93 districts under two different modeling frameworks, there are 558 estimates and we Bonferroni adjust the *p*-values for these. Assuming no further existing confounding effects on the sentence, we could interpret the estimated sentencing disparity (difference in estimated coefficients between two races) as the differential treatment due to race.

Results

National-level race-based sentencing disparities. We study 518,721 federal sentencing records from 2006 through 2020 (see “Methodology”). Figure 1 summarizes the results of linear regression models that account for the contribution of various factors to race-based disparities. Table 1 provides numerical values of these estimates. Within the federal system as a whole, the average sentence given to Black defendants is 18.5 months longer than that given to white ones. For Hispanic defendants, it is 5.3 months longer, and for defendants of another racial identity (ARI), it is 9 months shorter. Several comments are in order. Most importantly, these average differences are not conditioned at all: neither on type of crime, criminal history, or anything else.

Additionally, the grouping of all races other than Black, Hispanic, and white into the ARI category limits our understanding. For example, Asian defendants and Native American defendants are both in this group (called “other” in the government’s public data). For this reason, we limit our comments on results for ARI defendants in our nationwide analysis. We will return to a more detailed analysis of ARI defendants in our district-level analyses, in which the geography of each district hints at what groups might be driving sentencing disparities in that category.

Conditioning sentencing on personal demographics, namely age, gender (which the government measures as binary sex), and educational attainment, reduces the disparities to 12.9 months (−5.6) and 1.0 months (−4.3) for Black and Hispanic defendants respectively, demonstrating the role that these characteristics play in the disparate impact naively attributable to race. Conditioning on year of sentence has little effect, suggesting that time is not a primary driver of disparities, at least during the 2006–2020 time period spanned by our study. The presence of a guilty plea plays quite a modest role in explaining disparities for Hispanic defendants. In contrast, accounting for such pleas reduces the disparity for Black defendants from 12.8 to 10.3 months (−2.5) as compared to white ones.

Conditioning on criminal history further reduces the sentence differential for Black defendants to 2.2 months. This is a fairly large reduction, and it captures the extent to which the criminal histories of Black and white defendants differ. That is to say, the sentencing guidelines’ emphasis on criminal history accounts for a substantial proportion of the racial difference. In contrast, accounting for criminal history actually exacerbates the disparity between Hispanic and white defendants.

As one would expect based on how sentencing works (see “Introduction”), conditioning the actual sentence on presumptive sentence (labeled “grid cell” in Fig. 1 and Table 1) substantially reduces disparities for Black defendants, from 10.3 to 5.4 months (−4.9). The disparity for Hispanic defendants is no longer statistically significant. Additionally conditioning on mandatory minimum laws and government-sponsored downward departures reduces the Black-white disparity from 5.4 months to 1.9 months (−3.5, and compared with an unconditional difference of 18.5 months, −16.6) and yields a remaining Hispanic-white disparity, now statistically significant, of −1.1 months. This result means that controlling for all aforementioned factors, Hispanic defendants receive a conditional sentence that is slightly shorter than that given to white defendants. We caution that this result

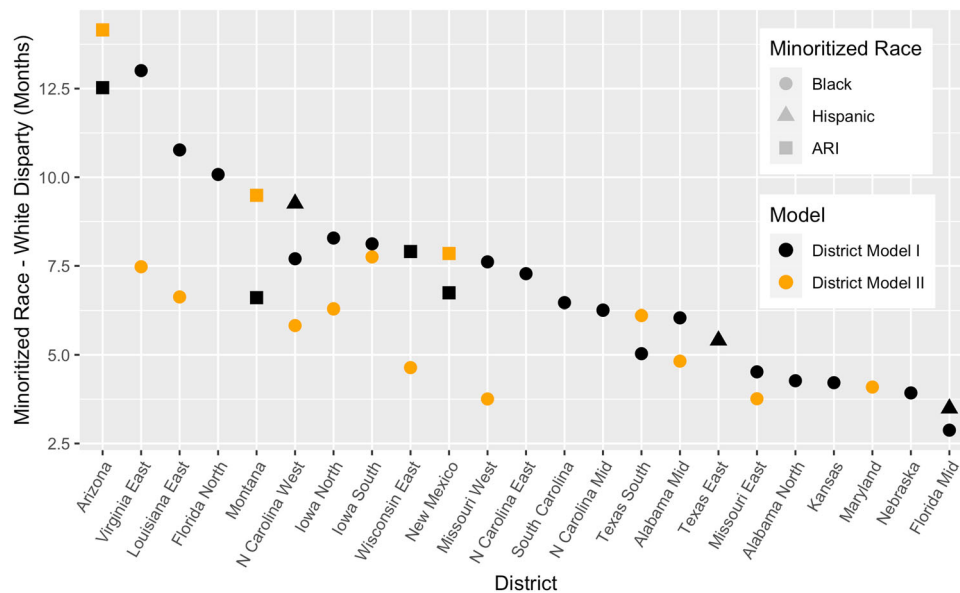


Fig. 2 Significant race-based sentencing disparities in federal judicial districts. We calculate the average sentencing disparity in months for defendants who are Black (circles), Hispanic (triangles), and another racial identity (ARI, squares), each as compared to white defendants. There are two sets of results. The first set comes from a regression model (District Model I, black symbols) that includes defendant demographics (age, sex, and educational attainment), sentencing year, presence of a guilty plea, relevant cell on the U.S. sentencing grid, presence of a mandatory minimum, presence of government-sponsored downward departures, and the interaction of judicial district and defendant race. In this model, all terms except for the interaction derive from nationwide data. For example, this model assumes that all districts, on average, apply the U.S. sentencing grid in the same way. For the second set of results (District Model II, orange symbols), we partition the same data by judicial district, and within each district perform a regression that includes the other aforementioned variables, as well as race. This model allows for variations in grid application, defendant demographics, and more. For each model, we present racial disparities unexplained by other factors that are statistically significant ($p < 0.05$ with Bonferroni adjustment). See Table 3 for estimates of the three racial disparities under each of the two modeling frameworks and see “Discussion” for the interpretation of these disparities vis-a-vis disparate impact and differential treatment of minoritized individuals. The analysis is based on $N = 518,721$ sentencing records. Model diagnostics, including F-statistics, the associated degrees of freedom, the associated p -values, as well as raw and adjusted r^2 values appear in our permanent data repository (Topaz, 2023). Each of these F-statistic p -values is numerically indistinguishable from zero. For District Model I, the adjusted r^2 value is 0.88. For District Model II, the adjusted r^2 values range from 0.70 to 0.91 with a median of 0.83.

does not mean that Hispanic defendants are exempt from disparities. For instance, when examined at a finer jurisdictional level, disparities for Hispanic defendants are well-established (Ulmer and Parker, 2020).

Finally, to begin accounting for the federal judicial district where sentencing occurs, we include the district as a main effect in our model. That said, the observed effect of placing this control is modest. To achieve a more refined understanding, we will consider the role of the district in much greater detail below.

After conditioning the sentence outcome on age, gender, educational attainment, sentencing year, presence of a guilty plea, presumptive sentence, application of mandatory minimum laws, government-sponsored downward departures, and judicial district, there remains a statistically significant Black-white disparity of 1.9 months. The Hispanic-white disparity is negligible and not statistically significant. The ARI-white disparity of 2.8 months is significant and yet is difficult to interpret, as mentioned previously, due to the government combining many different racial groups into one. In summary, even after controlling for numerous factors that should explain sentencing disparities, there remain unexplained differences for Black and ARI defendants. To investigate the loci of these differences, we will proceed to an examination of each individual judicial district. We would like to re-emphasize that while our model and assumptions lead to no unexplained disparity for Hispanic defendants at the national level, prior results (see “Introduction”) suggest that a disparity does exist, and thus whether certain districts display it is a question that merits investigation, just as it does for Black and ARI defendants.

Race-based sentencing disparities within districts. To study associations between sentence length and defendant race within districts, we use two models. Figure 2 and Table 3 display the results of these models. The first model, labeled District Model I, measures the still-unexplained racial disparity within each district by including the interaction of district and defendant race while accounting for a nationwide-aggregated age, gender, educational attainment, sentencing year, presence of a guilty plea, presumptive sentence, mandatory minimum laws, and government-sponsored downward departures. The second model, labeled District Model II, allows for interdistrict variations in defendant demographics, in the application of presumptive sentences, and more. Arguably, the most important role of Model I is to set a benchmark for comparison to Model II, allowing us to ask how racial disparities change once accounting for structural factors such as in-district demographics and sentencing practices. See “Methodology” as well as the captions of Fig. 2 and Table 3 for further explanation.

In District Model I, 22 districts have statistically significant unexplained disparities for one or more minoritized racial groups ($p < 0.05$, Bonferroni-adjusted). Seventeen districts have a Black-white disparity, ranging from a high of 13.0 ± 1.6 months (95% confidence interval) for the Eastern District of Virginia to 2.9 ± 1.4 months for the Middle District of Florida. Three districts have a Hispanic-white disparity, ranging from 9.3 ± 4.6 months for the Western District of North Carolina to 3.5 ± 1.8 months for the Middle District of Florida. Finally, four districts have an ARI-white disparity, ranging from 12.5 ± 1.6 months for the District of Arizona to 6.6 ± 2.1 months for the District of Montana.

Table 3 Race-based sentencing disparity in the federal court system.

District	Black-white disparity (mo.)		Hispanic-white disparity (mo.)		ARI-white disparity (mo.)	
	Model I	Model II	Model I	Model II	Model I	Model II
ALM	6.0 ± 2.4***	4.8 ± 2.3*				
ALN	4.3 ± 1.8**					
AZ					12.5 ± 1.6****	14.2 ± 1.6****
FLM	2.9 ± 1.4*		3.5 ± 1.8*			
FLN	10.1 ± 3.4****					
IAN	8.3 ± 3.1****	6.3 ± 3.0*				
IAS	8.1 ± 2.8****	7.8 ± 2.8****				
KS	4.2 ± 2.0**					
LAE	10.8 ± 2.3****	6.6 ± 2.4****				
MD		4.1 ± 2.0*				
MOE	4.5 ± 1.1****	3.8 ± 1.1****				
MOW	7.6 ± 1.5****	3.8 ± 1.5***				
MT					6.6 ± 2.1****	9.5 ± 2.2****
NCE	7.3 ± 2.0****					
NCM	6.3 ± 1.8****					
NCW	7.7 ± 1.8****	5.8 ± 1.9****	9.3 ± 4.6*			
NE	3.9 ± 2.0*					
NM					6.7 ± 2.8***	7.8 ± 2.7****
SC	6.5 ± 1.4****					
TXE			5.4 ± 1.9****			
TXS	5.0 ± 1.9****	6.1 ± 1.8****				
VAE	13.0 ± 1.6****	7.5 ± 1.7****				
WIE		4.6 ± 2.2*			7.9 ± 4.0*	

Here we provide the numerical values underlying Figs. 2-4. Each entry represents mean sentencing disparity (in months, along with a 95% confidence interval calculated using standard errors) for a federal judicial district as estimated from a regression model. District Model I includes defendant demographics (age, sex, and educational attainment), sentencing year, presence of a guilty plea, relevant cell on the U.S. sentencing grid, presence of a mandatory minimum, presence of government-sponsored downward departures, and the interaction of judicial district and defendant race. In this model, all terms except for the interaction derive from nationwide data. For example, this model assumes that all districts, on average, apply the U.S. sentencing grid in the same way. In contrast, for District Model II, we partition the same data by judicial district, and within each district perform a regression that includes the other aforementioned variables, as well as race. This model allows for variations in grid application, defendant demographics, and more. Asterisks denote the range for a Bonferroni-adjusted *p*-value: * (*p* < 0.05), ** (*p* < 0.01), *** (*p* < 0.001), **** (*p* < 0.0001). Table cells missing an entry indicate no statistically significant disparity (at a level of 0.05), as do districts absent from the table. Two-letter district abbreviations are the common postal abbreviations of states, and apply when a district contains its entire state. Multidistrict states code each district with three-letter abbreviations, where the third letter indicates a geographic region: C = Central, E = East, M = Middle, N = North, S = South, W = West. Sample sizes for the listed districts are ALM (2502), ALN (5121), AZ (12,687), FLM (11,981), FLN (3637), IAN (3677), IAS (4276), KS (5565), LAE (3727), MD (7965), MOE (10,360), MOW (9225), MT (4391), NCE (7,1401), NCM (4627), NCW (6367), NE (4961), NM (7245), SC (9775), TXE (8792), TXS (16,409), VAE (10,405), WIE (4280). Model diagnostics, including *F*-statistics, the associated degrees of freedom, the associated *p*-values, as well as raw and adjusted *r*² values appear in our permanent data repository (Topaz, 2023). Each of these *F*-statistic *p*-values is numerically indistinguishable from zero. For District Model I, the adjusted *r*² value is 0.88. For District Model II, the adjusted *r*² values range from 0.70 to 0.91 with a median of 0.83.

In District Model II, 14 districts have statistically significant unexplained disparities (*p* < 0.05, Bonferroni-adjusted), each for exactly one minoritized racial group. Eleven districts have a Black-white disparity, ranging from a high of 7.8 ± 2.8 months for the Southern District of Iowa to 3.8 ± 1.1 months for the Eastern District of Missouri and 3.8 ± 1.5 months for the Western District of Missouri. Three districts have an ARI-white disparity, ranging from 14.2 ± 1.6 months for the District of Arizona to 7.8 ± 2.7 months for the District of New Mexico.

Discussion

As shown in Fig. 1, the system-wide raw disparities for Black and Hispanic defendants decrease substantially once conditioned on defendant-specific and case-specific characteristics. This result is consistent with prior studies (Light, 2022; Mitchell, 2005; Rehavi and Starr, 2014; Spohn, 2013, 2000). For Black defendants, personal demographics (age, sex, educational attainment) and criminal history appear to play the most significant roles. In contrast, a raw disparity in favor of ARI defendants flips its sign after conditioning. However, interpreting this result is difficult as the actual racial identification of individuals in this group is unknown. We do not observe a significant unexplained system-wide disparity for the conditional sentences of Hispanic defendants. Much of the disparity appears attributable to personal demographics, as was the case for Black defendants.

Of course, the federal judicial system is not centralized. Sentences are created as part of a process by a courtroom workgroup within individual districts, and thus a more granular examination is warranted. Indeed, a meta-analysis of sentencing disparities research notes “[results] that pool data from many jurisdictions... may suffer from

aggregation bias,” (Mitchell, 2005). We performed two analyses to assess the contribution of the federal judicial district to sentencing disparities. District Model I assumes that the sentencing guidelines are implemented uniformly across all districts, while the framework of District Model II conditions on all factors, including the application of the sentencing guidelines, uniquely for each district. As mentioned previously, the most important role of District Model I is to set a benchmark for comparison to District Model II. The two models inform our understanding of the disparate impact and differential treatment in the following manner.

First, for 24 district-race combinations, District Model I shows a statistically significant disparity (black symbols in Fig. 2). Specifically, 17 districts show a disparity for Black individuals, three for Hispanic individuals, and four for ARI individuals. Within the framework set by this model, these results have at least two interpretations. One interpretation is that there may be bias of courtroom actors, that is, differential treatment, for the identified district-race combinations. We will consider this point in more detail momentarily. The second interpretation is that the system has a structural disparate impact. As a concrete example, the estimated Black-white disparity of 13.0 months in the Eastern District of Virginia could indicate that for socioeconomic reasons, Black defendants are more strongly associated with offenses and criminal histories that have longer presumptive sentences. To address this type of structural disparity requires one to think not only about socioeconomic determinants, but also about who is policed, who is arrested, who is prosecuted, who is convicted, and why particular cells in the sentencing grid carry longer presumptive penalties. In short, the elimination of these disparities would require structural change focused outside of the courtroom.

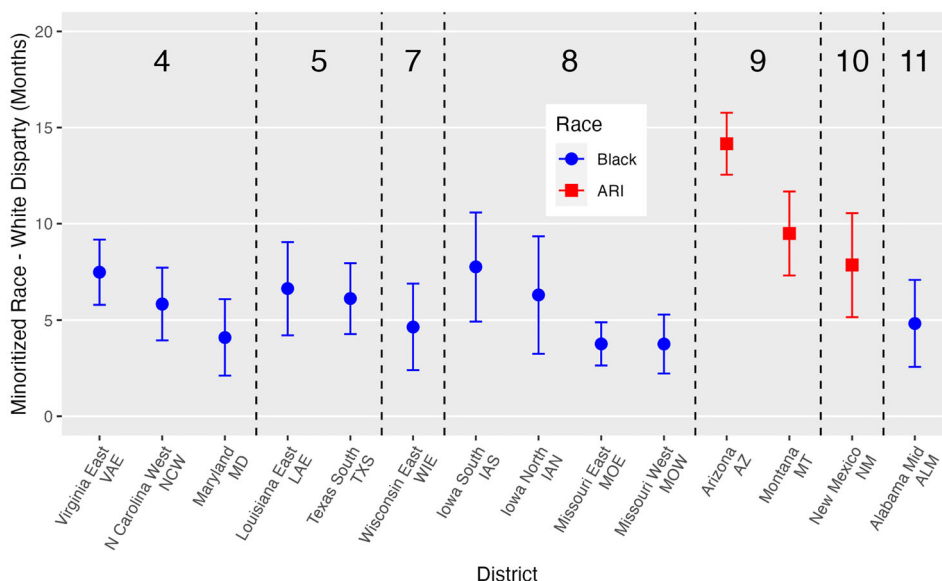


Fig. 3 Summary of evidence of race-based criminal sentencing bias (differential treatment) in federal judicial districts. In Fig. 2, 14 federal districts display evidence of differential treatment by judges and other members of the courtroom workgroup even in a modeling framework that allows sentences to be conditioned separately for each district. In the figure above, we summarize the disparities and visualize their 95% confidence intervals. Eleven districts show evidence of bias against Black defendants (blue circles). Three districts show evidence of bias against defendants who have another racial identification (ARI, red squares), a category that includes Native Americans. The fourteen districts come from seven of the federal circuits. Circuit numbers appear in the center of each subdivided section above.

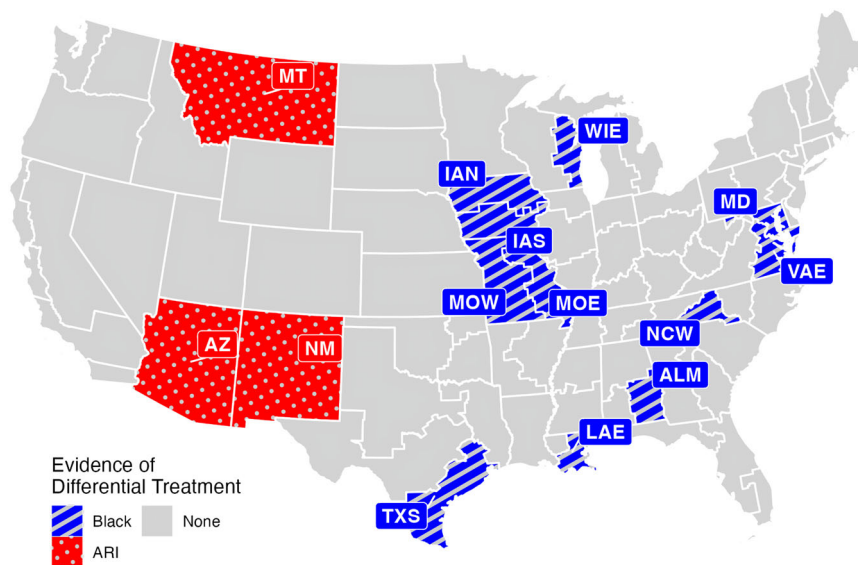


Fig. 4 Map of evidence of race-based criminal sentencing bias (differential treatment) in federal judicial districts. This map puts Fig. 3 and Table 3 in geographic context; see those captions for further explanation. The horizontal axis labels in Fig. 3 provide the full district name corresponding to each abbreviation in the map. States marked “none” are those for which our study did not produce evidence of differential treatment. The designation “none” does not mean that differential treatment may not, in reality, exist.

Second, for 20 district-race combinations, District Model I shows a statistically significant disparity (black symbols in Fig. 2) that is either reduced or lost in District Model II (orange symbols). To restate this result, conditioning on all factors at the individual district-level results in a reduced estimate of racial disparity for these 20 cases. We interpret this reduction in racial disparity as evidence of differential impact on minoritized groups. For example, in the Eastern District of Virginia, the disparity of 13.0 months for Black individuals in District Model I is reduced

to 7.5 months in District Model II, suggesting that in-district factors—including the specific ways in which that district implements the sentencing grid—explain 5.5 months of the disparity measured in District Model I. As shown in Fig. 2 and Table 3, the 20 aforementioned cases consist of 16 districts with disparities for Black individuals, three districts for Hispanic individuals, and one district for ARI individuals. Policies encouraging more consistent application of the guidelines could help reduce such disparities. Overall, our results about differential impact

extend previous work that, irrespective of race, identifies substantial inter-jurisdictional discrepancies in sentencing outcomes within state systems (Ridgeway et al., 2020; Wright, 2012) and the federal system (Harries and Lura, 1973; United States Sentencing Commission, 2020a; Wu and Spohn, 2010).

Finally, for 14 district-race combinations, District Model II shows a significant disparity (regardless of results from District Model I). Another way to state this result is that, for these 14 cases, even after conditioning on all factors at the district level, an unexplained racial disparity remains. Figure 3 summarizes these 14 cases, provides confidence intervals for the disparities, and organizes them by federal circuit. There are Black-white disparities in the Fourth, Fifth, Seventh, Eighth, and Eleventh Circuits, and ARI-white disparities in the Ninth and Tenth Circuits. Figure 4 provides a map of these disparities. The Black-white disparities are largely concentrated in the plains, the mid-Atlantic seaboard, and scattered across the deep south. The ARI-white disparities are in Montana, New Mexico, and Arizona, which are states with high per capita Native American populations (United States Census Bureau, 2021). However, due to the racial heterogeneity of the ARI category, and due to complex issues of jurisdiction on tribal lands in the United States, our study cannot provide a definitive and nuanced interpretation of the ARI-white disparity.

The results summarized in Figs. 3 and 4 are racial disparities that are explained neither by defendant demographics (age, sex, educational attainment), sentencing year, the presence of a guilty plea, presumptive sentence (based on criminal history and severity of the offense), mandatory minimum laws, or government-sponsored downward departures. We interpret these unexplained disparities as evidence of possible differential treatment of defendants, whether due to implicit or explicit bias of judges, prosecutors, and defense attorneys. One study that focused on federal drug sentencing described this type of differential treatment as “state-level racial animus” (Tuttle, 2019).

Conclusions

While racial disparities and interdistrict variability in sentencing have both been previously established in the literature, their interplay has not. We have provided a look at district-level variations in race-based sentencing disparity and identified 14 judicial districts showing significant differences for minoritized defendants as compared to white ones.

It is important to keep in mind several limitations of our work:

1. We have focused on sentence length as opposed to the judge’s decision of whether or not to impose prison time at all. Though not the subject of our study, the imprisonment decision itself is an important source of racial disparity (Ulmer et al., 2016).
2. We have not disaggregated our results by offense type. Offense type is not central to our research question, though it is possible that in some districts, disparities exist for some offense types and not others. This could be an appropriate route of investigation for future work.
3. As we mentioned previously, and in keeping with other published work, we have omitted noncitizens from our analysis. Unfortunately, this likely limits the understanding of disparities experienced by Hispanic individuals.
4. Evidence of differential treatment within a district is not proof and, as with any statistical modeling study, hinges on the modeling assumptions we have articulated (see “Methodology”).
5. Equally, evidence of differential treatment does not reveal which judges, prosecutors, or defense attorneys, if any, might be responsible.

6. On the other hand, there could certainly be differential treatment by judges, prosecutors, and defense attorneys in districts other than those we have identified here. Biased sentences in those districts could remain undetected in our analysis due to other sentences that mask or counterbalance them.

While the actions of members of courtroom workgroups are certainly not the only source of inequity in the federal court system—as our results confirm—their decisions should be subject to close scrutiny and analysis. In the 1980 case *Richmond Newspapers v. Virginia* (448 U.S. 55), the Supreme Court found that openness of criminal courtrooms is “implicit in the guarantees of the First Amendment.” This means that in theory, the public is allowed to know the sentence given by each judge to each defendant. However, that theoretical right does not translate into public transparency on a large scale because USSC does not include the names of the members of the courtroom workgroup who created the sentence in its public data, thereby precluding analysis at the individual level. Some preliminary steps have been taken to unmask judge identities, including the JUSTFAIR database of approximately 600,000 federal sentences (Ciocanel et al., 2020). Of course, any such large-scale, data science-based effort to identify sentencing judges is subject to error. For this reason, in order to enable the most accurate, and ideally, causal, analyses of race-based differential treatment, the government would need to cease redacting judge identity in its public records and find ways to identify other actors in the courtroom workgroup who help shape the outcome.

Data availability

Raw data come from the Individual Offender Files of the United States Sentencing Commission, available publicly at <https://www.ussc.gov/research/datafiles/commission-datafiles>. Our derived dataset and model diagnostics appear in our permanent data repository Topaz (2023).

Received: 10 December 2022; Accepted: 7 June 2023;

Published online: 29 June 2023

References

- Abrams DS, Bertrand M, Mullainathan S (2012) Do judges vary in their treatment of race? *J Legal Stud* 41(2):347–383
- Arnold D, Dobbie W, Yang CS (2018) Racial bias in bail decisions. *Q J Econ* 133(4):1885–1932
- Baumer EP (2013) Reassessing and redirecting research on race and sentencing. *Justice Q* 30(2):231–261
- Bushway SD, Forst B (2013) Studying discretion in the processes that generate criminal justice sanctions. *Justice Q* 30(2):199–222
- Bushway SD, Piehl AM (2001) Judging judicial discretion: legal factors and racial discrimination in sentencing. *Law Soc Rev* 35(4):733
- Bushway SD, Piehl AM (2011) Location, location, location: the impact of guideline grid location on the value of sentencing enhancements. *J Empir Legal Stud* 8:222–238
- Ciocanel MV, Topaz CM, Santorella R et al (2020) JUSTFAIR: judicial system transparency through federal archive inferred records. *PLoS ONE* 15(10):e0241381
- Dixon TL (2006) Schemas as average conceptions: skin tone, television news exposure, and culpability judgments. *Journal Mass Commun Q* 83(1):131–149
- Dixon TL, Linz D (2000) Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *J Comm* 50(2):131–154
- Doerner JK, Demuth S (2010) The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in U.S. federal courts. *Justice Q* 27(1):1–27
- Feldmeyer B, Ulmer JT (2011) Racial/ethnic threat and federal sentencing. *J Res Crime Delinq* 48(2):238–270

- Franklin TW (2013) Sentencing Native Americans in US federal courts: an examination of disparity. *Justice Q* 30(2):310–339
- Frase RS, Roberts JV (2020) *Paying for the past: the case against prior record sentence enhancements*. Oxford University Press, New York
- Harries KD, Lura R (1973) The geography of justice-sentencing variations in U.S. judicial districts. *Judicature* 57:392
- Johnson BD, Ulmer JT, Kramer JH (2008) The social context of guidelines circumvention: the case of federal district courts. *Criminology* 46(3):737–783
- King RD, Light MT (2019) Have racial and ethnic disparities in sentencing declined? *Crime Just* 48(1):365–437
- Kramer J, Steffensmeir D (1993) Race and imprisonment decisions. *Social Q* 34(2):357–376
- Light MT (2022) The declining significance of race in criminal sentencing: evidence from US federal courts. *Soc Forces* 100(3):1110–1141
- Mitchell O (2005) A meta-analysis of race and sentencing research: explaining the inconsistencies. *J Quant Criminol* 21(4):439–466
- Mustard DB (2001) Racial, ethnic, and gender disparities in sentencing: evidence from the US federal courts. *J Law Econ* 44(1):285–314
- Rachlinski JJ, Wistrich AJ (2017) Judging the judiciary by the numbers: empirical research on judges. *Ann Rev Law Soc Sci* 13:203–229
- Rehavi MM, Starr SB (2014) Racial disparity in federal criminal sentences. *J Pol Econ* 122(6):1320–1354
- Ridgeway G, Moyer RA, Bushway SD (2020) Sentencing scorecards: reducing racial disparities in prison sentences at their source. *Criminol Public Pol* 19(4):1113–1138
- Scott RW (2010) Inter-judge sentencing disparity after Booker: a first look. *Stanford Law Rev* 63(1):1–66
- Smith RJ, Levinson JD, Robinson Z (2014) Implicit white favoritism in the criminal justice system. *Ala L Rev* 66:871–923
- Spohn C (2013) The effects of the offender's race, ethnicity, and sex on federal sentencing outcomes in the guidelines era. *Law Contemp Prob* 76(1):75–104
- Spohn CC (2000) Thirty years of sentencing reform: the quest for a racially neutral sentencing process. In: Horney J (ed) *Policies, processes, and decisions of the criminal justice system*, vol 3. United States Department of Justice, Washington, D.C., pp 427–502
- Starr S (2013) Did Booker increase disparity? Why the evidence is unpersuasive. *Fed Sent Rep* 25(5):323–326
- Steffensmeir D, Ulmer J, Kramer J (1998) The interaction of race, gender, and age in criminal sentencing: the punishment cost of being young, Black, and male. *Criminology* 36(4):763–798
- Topaz CM (2023) Data for Federal criminal sentencing: race-based disparate impact and differential treatment in judicial districts. *figshare* https://figshare.com/articles/journal_contribution/Federal_Criminal_Sentencing_Race-Based_Disparate_Impact_and_Differential_Treatment_in_Judicial_Districts/22083350
- Tuttle, C. Racial disparities in federal sentencing: evidence from drug mandatory minimums. <https://doi.org/10.2139/ssrn.3080463> (2019)
- Ulmer J, Painter-Davis N, Tinik L (2016) Disproportional imprisonment of Black and Hispanic males: sentencing discretion, processing outcomes, and policy structures. *Justice Q* 33(4):642–681
- Ulmer JT, Parker BR (2020) Federal sentencing of Hispanic defendants in changing immigrant destinations. *Justice Q* 37(3):541–570
- United States Census Bureau (2021) Race and ethnicity in the United States: 2010 census and 2020 census. <https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html>
- United States Sentencing Commission (1987) 1987 Supplementary report on the initial sentencing guidelines and policy statements. <https://www.ussc.gov/guidelines/guidelines-archive/1987-supplementary-report-initial-sentencing-guidelines-and-policy-statements>
- United States Sentencing Commission (2018) Demographic differences in sentencing: an update to the 2012 Booker report. *Fed Sent Rep* 30(3):212–229
- United States Sentencing Commission (2020a) Inter-district differences in federal sentencing practices. <https://www.ussc.gov/research/research-reports/inter-district-differences-federal-sentencing-practices>
- United States Sentencing Commission (2020b) Variable codebook for individual offenders. https://www.ussc.gov/sites/default/files/pdf/research-and-publications/datafiles/USSC_Public_Release_Codebook_FY99_FY20.pdf
- United States Sentencing Commission (2021) Individual offender data files. <https://www.ussc.gov/research/datafiles/commission-datafiles>
- Vagins DJ, McCurdy J (2006) Cracks in the system: twenty years of the unjust federal crack cocaine law. Technical Report. American Civil Liberties Union, Washington DC
- Wright RF (2012) Persistent localism in the prosecutor services of North Carolina. *Crime Justice* 41(1):211–264
- Wu J, Spohn C (2010) Interdistrict disparity in sentencing in three U.S. district courts. *Crime Delinq* 56(2):290–322
- Yang CS (2014) Have interjudge sentencing disparities increased in an advisory guidelines regime? Evidence from Booker. *N Y Univ Law Rev* 89:1268–1342
- Yang CS (2015) Free at last? Judicial discretion and racial disparities in federal sentencing. *J Legal Stud* 44(1):75–111

Acknowledgements

We are grateful to Nicholas Goldrosen and Christian M. Smith for their contributions during the formative stages of this work.

Author contributions

All authors made substantive contributions to the design of the study, the analysis and/or interpretation of the data, and the drafting and revision of the manuscript. All authors have given final approval to the published version of this work, and agree to be accountable for it.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Chad M. Topaz.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023