

2013

Did Booker Increase Sentencing Disparity? Why the Evidence Is Unpersuasive

Sonja B. Starr

University of Michigan Law School, sbstarr@umich.edu

Available at: <https://repository.law.umich.edu/articles/1192>

Follow this and additional works at: <https://repository.law.umich.edu/articles>



Part of the [Criminal Law Commons](#), [Criminal Procedure Commons](#), and the [Law and Race Commons](#)

Recommended Citation

Starr, Sonja B. "Did Booker Increase Sentencing Disparity? Why the Evidence Is Unpersuasive." Fed. Sent'g Rep. 25, no. 5 (2013): 323-6.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

Did *Booker* Increase Disparity? Why the Evidence Is Unpersuasive



SONJA STARR*

Assistant Professor,
University of
Michigan Law
School

The Sentencing Commission's recent report on the effects of *United States v. Booker* makes a number of very worrisome claims. The most alarming is that the gap in sentences between otherwise similar Black and White men has nearly quadrupled: from 4.5 percent before *Booker*, to 15 percent after it, to 19.5 percent after *United States v. Kimbrough* and *United States v. Gall*.¹ The Commission further claims that interjudge disparity has increased in two-thirds of the federal districts, and that interdistrict variation has also increased.² If its findings were accurate, and if these changes could be causally attributed to *Booker* and its successors, it would clearly raise very serious policy concerns. The Commission evidently believes that *Booker* has in fact created a problem, because it concludes by proposing changes that would essentially undo *Booker* and restore presumptive guidelines.

Fortunately, however, the Commission's conclusions are unsupported. The best available evidence suggests that racial disparity in the federal criminal justice system is a persistent problem, but has not worsened after *Booker* (much less as a result of *Booker*). There are a number of problems with the Commission's approach, but the most important are (1) its misleading method of measuring disparity and (2) its failure to use methods that can support valid causal inferences about *Booker's* effects. Here, I describe each of these problems and summarize the key findings of my own recent study of *Booker's* effects, coauthored with economist Marit Rehavi, which was designed to avoid these problems and which reaches conclusions contrary to the Commission's. The study, which includes a more detailed critique of the Commission's methods, is forthcoming in the *Yale Law Journal*.³ Our study and my discussion here focus on race, but the methodological concerns also undermine the Commission's findings on interjudge and interdistrict disparities.

I. The Commission's Misleading Measure of Disparity

In any study that seeks to isolate the legally unwarranted components of sentence disparity, researchers must try to make sure that the cases they are comparing are similar. Ideally, researchers would want to have perfect information about the defendant's true underlying criminal conduct and control for it, but of course getting such information is impossible. But researchers can at least identify the component of disparity that is unexplained by observable,

legally relevant variables. They usually employ some form of *regression*, a statistical tool that is useful for estimating the relationship between a particular variable (here, race) and the outcome (sentence length) when other "control" variables (for instance, crime characteristics and criminal history) are held constant. The Commission emphasizes the strengths of this approach, but there is nothing inherently trustworthy about regression. There are many ways for regression to produce misleading estimates, starting with the omission of appropriate control variables or the inclusion of inappropriate ones. In the Commission's regression model, three key controls are the "presumptive sentence" (the bottom end of the Guidelines range), the statutory minimum, and "departure status" (whether and in what direction the judge departed).⁴ All three are deeply problematic.

Consider, first, the presumptive sentence. Controlling for this variable means that the Commission is in essence asking whether Blacks and Whites in the same cell of the Guidelines grid received the same sentences. But the Guidelines cell is not a simple proxy for underlying conduct and criminal history. It results from discretionary processes—charging, plea bargaining, and fact-finding—which might be just as susceptible to racial disparity as the final sentencing decision is. If so, then Blacks and Whites with the same underlying conduct would end up in different Guidelines cells, so comparing blacks and whites in the *same* Guidelines cell would be apples-to-oranges. Suppose, for instance, that Blacks and Whites averaged the same actual sentences *relative to their presumptive sentences*, but that there was serious disparity in fact-finding, such that Blacks received offense levels averaging six levels higher than Whites with the same underlying conduct. In this hypothetical, defendants with the same underlying crimes and criminal histories would end up with sharply different sentences depending on their race, yet the Commission's method would misleadingly estimate zero racial disparity.⁵

The same concern arises with respect to the mandatory minimum, and there is good reason to believe that mandatory minimums are an important source of racial disparity in sentences. In another recent study, Prof. Rehavi and I found that Black male federal defendants were nearly *twice* as likely as White men to be charged with a mandatory minimum offense, after controlling for the arrest offense, criminal history, and other prior characteristics; this

Federal Sentencing Reporter, Vol. 25, No. 5, pp. 323–326, ISSN 1053-9867, electronic ISSN 1533-8363.
© 2013 Vera Institute of Justice. All rights reserved. Please direct requests for permission to photocopy or reproduce article content through the University of California Press's Rights and Permissions website, <http://www.ucpressjournals.com/reprintInfo.asp>. DOI: 10.1525/fsr.2013.25.5.323.

charging gap appears to explain more than half of the otherwise unexplained Black–White sentence gap.⁶ Yet the Commission’s method filters out the share of sentence disparity that is driven by mandatory minimums. More generally, it filters out any contributions of prosecutorial decisions to sentence disparities, even though the dominant role of prosecutors in shaping federal sentences is widely recognized.⁷

Even if for some reason one thought it sensible to focus exclusively on judges’ decisions, the Commission’s method does not even capture all aspects of the judicial sentencing decision. As noted above, its estimates exclude disparities in sentencing fact-finding. Moreover, by including the departure status control, the Commission also filters out any disparities in departure rates and direction—a bewildering decision. What the Commission is left estimating is, essentially, disparities in departure magnitude (and choice of sentences within the Guidelines range). But the average departure that Black or White defendants receive depends on the rate, direction, and magnitudes of departures taken together; trying to analyze magnitude while controlling for rate and direction is nonsensical.⁸

In fairness to the Commission, the practice of controlling for the presumptive sentence and mandatory minimum is common in the sentencing literature (though the addition of the departure status control is the Commission’s own unfortunate innovation).⁹ The reason for this is probably data limitations. Sentencing data sets, most notably the Commission’s, are typically limited to the sentencing stage, making it hard to assess disparities emerging from earlier stages. But the fact that a practice is common or convenient does not make it correct, and researchers have not advanced a justification for this approach other than necessity. Fortunately, an alternative approach is now possible, as explained in Part III.

When designing a regression model, it is important to have a clear sense of exactly what one is trying to measure—here, we want to know what kinds of “sentencing disparity” are important for normative and policy purposes. Although this inquiry is complex, one core principle that seems uncontroversial is that policymakers should aim for equal treatment by the *whole criminal justice system*—not just for equalizing sentences relative to the Guidelines cell and the mandatory minimum.¹⁰ And if that is policymakers’ objective, they should not make decisions based on empirical studies that assess the sentencing decision in isolation.

The Commission’s narrow approach risks counterproductive policy recommendations. If one conceives of “sentencing disparity” strictly in terms of departures from the Sentencing Guidelines, then the solutions one considers will likely focus on constraining those departures—as the Commission’s proposal does. But there is good reason to believe such constraints might actually increase the sentence gap between Blacks and Whites with the same underlying criminal conduct. Constraining judicial discretion, historically, strengthens the power of prosecutors,

whose charging and bargaining decisions more directly determine the sentence.¹¹ And if (as our research suggests) prosecutorial decisions are a key source of disparity, shifting power from judges to prosecutors might make disparity worse. In fact, we found that the Black–White sentence gap was heavily driven by the cases in which judges have the *least* sentencing discretion—those with mandatory minimums. Likewise, it is premature to turn to tightened sentencing Guidelines in response to interdistrict and interjudge disparities when there are no empirical studies as to whether *prosecutors’* decisions also vary—within or between districts—in ways that create sentence disparities. Indeed, the Commission itself cites anecdotal evidence that they do vary, but then ignores that evidence’s implications for its policy proposals.¹²

II. Causal Inference about *Booker’s* Effects

Even if it were measuring disparity correctly, the Commission also provides no basis for concluding that *Booker*, or its progeny *Kimbrough* and *Gall*, actually *caused* the increase in unwarranted sentence disparity that purportedly occurred in recent years. Causal inference based on changes over time is a fraught enterprise because many things change over time. Racial disparity might grow over a period of years for many reasons. For instance, changes in the underlying case mix, changes in DOJ priorities, or changes in the composition of the bench or U.S. attorneys’ offices might all have racially disparate effects. The Commission’s method provides no way to distinguish *Booker’s* effects from background trends or unrelated events.

Instead, the Commission simply conducts separate regressions on four different samples, divided by sentencing date: (1) from *United States v. Koon* (1996) to the PROTECT Act (2003), (2) between PROTECT and *Booker*, (3) between *Booker* and *Kimbrough/Gall*, and (4) after *Kimbrough/Gall*. The Commission then simply compares the estimates from each period of the average effect of being Black on the sentence. It finds the smallest effect in period 2 (when the Guidelines were at their tightest), a big rise in period 3, and a further rise in period 4. One basic problem with this approach is the Commission’s unconventional choice to run separate regressions for each time period, which means that the effects of the various control variables are not held constant across time periods.

More fundamentally, a finding that racial disparity was larger after *Booker* does not mean it got larger *because* of *Booker*. Suppose, for instance, that racial disparity increased gradually at a constant rate from 2003 (when the PROTECT Act passed) to 2011 (the end of the study period). If this were true, it could entirely explain the Commission’s results: the later time periods would have larger average disparities. But if this continuous background trend was not altered when *Booker* and *Kimbrough/Gall* were decided, there is no logical reason to think the cases had any effect on disparity. In fact, even if *Booker* and *Kimbrough/Gall* had *slowed* the rate of increase in disparity, the average would *still* be higher in the later periods. In that case, the

Commission's method would suggest that the decisions increased disparity even though they actually had the opposite effect.

In other words, the Commission's method tells us nothing about *Booker's* causal effects. This is true not just for its racial disparity analysis, but also for its analyses of other disparities. And yet the Commission's policy recommendations assume that *Booker* created a problem that needs to be fixed.¹³

III. An Alternative Assessment of *Booker's* Effects on Disparity

In our forthcoming article, Marit Rehavi and I present an alternative analysis of *Booker's* effects on racial disparity that reaches very different results.¹⁴ Using linking files from the Bureau of Justice Statistics, we constructed a multi-agency data set that traces cases from arrest through sentencing, so that we could assess disparities throughout the post-arrest justice process. Our sample consisted of non-immigration cases involving Black and White defendants from 2003 to 2009—the period in which the Commission found the most dramatic rise.¹⁵

Instead of the presumptive sentence, mandatory minimum, and/or departure status, our key crime-severity control was the *arrest offense*. The Marshals' Service records arrest-offense information in detail (430 distinct codes), which means this is a much richer offense-type control than the broad categories (e.g., "drugs") that the Commission uses. We further supplemented it with additional arrest offense-related details from other fields. Other control variables included criminal history category, education, age, gender, district, and U.S. citizenship. Controlling for the arrest offense means that our disparity estimates do not capture disparity in the *arrest* process, but they at least capture the aggregate sentence disparity produced from charging through sentencing—a much broader swath of the criminal process than that covered by the Commission's estimates.¹⁶ In short, we considered whether otherwise similar Black and White defendants who were arrested for the same crimes ultimately ended up with different sentences, and whether this race gap changed over time. We also conducted additional analyses to assess disparities (and changes in disparities) in particular stages such as charging.

We find that Black arrestees do face significantly longer sentences compared to similar White arrestees—but the good news is that there is no apparent increase in disparity over the period 2003 to 2009. Indeed, we actually find a decline in disparity, though not a significant one.¹⁷ The Commission's finding of an increase over that period is an artifact of its misleading approach to defining disparity, excluding the components of the sentence gap that are driven by unexplained disparities in charging, plea bargaining, fact-finding, and departure rates.

These long-term trends (or lack thereof) do not tell us anything about *Booker's* causal effects, but we conducted additional analysis to get at the causal question, using

a rigorous method called regression discontinuity design (RDD). This method filters out continuous background trends in racial disparity, assessing whether there are sudden *breaks* in those trends just after *Booker*. If we assume *Booker* is the only important *sudden* change occurring right at that time, then such trend breaks can probably be attributed to the decision. The central limitation of RDD is that it does not assess *Booker's* long-term effects; unfortunately, there is no good way to analyze those effects rigorously. However, focusing on the short-term effects can be informative. After all, *Booker* had a very sudden effect on departure rates; they jumped by nearly ten percentage points in the first month, a sharp break in an otherwise smooth trend.¹⁸ If judges were inclined to use their newfound discretion in ways that increased racial disparity, one would expect at least some of that effect to be seen right away, because judges began to exercise that discretion immediately.

Using this method, we find no evidence that *Booker* increased aggregate sentence disparity—in fact, we find a small but statistically significant *reduction* of disparity in *Booker's* immediate wake.¹⁹ Nor do we find any notable racially disparate changes in charging, charge bargaining, fact-finding, or departures.²⁰ Indeed, the only notable effect of *Booker* is the sudden increase in departures, but that increase appears to have benefited Black and White defendants equally.

IV. Conclusion

The Commission's report is alarming, but the alarm is unfounded. To date, there is no convincing evidence that the Black–White sentence gap is larger after *Booker* or its progeny, and certainly no evidence that the gap grew *because of Booker*.²¹ Moreover, the report's methodological flaws are also applicable to its findings concerning inter-district and interjudge disparities. There is no reason to consider the Commission's "solutions" at this point, because there is no evidence that *Booker* has created a problem.

The Commission recommends requiring substantial weight to be given to the Guidelines, greater justification for large departures, heightened appellate review of policy-based departures, and appellate deference to Guidelines sentences. These proposals are in tension with *Booker*, and especially with *Kimbrough*, *Gall*, and *United States v. Rita*,²² and their constitutionality is therefore unclear, but I leave that point for others to explore. Regardless, the evidence suggests the proposals would be unwise policy: tightening judicial discretion seems, if anything, likely to be a counterproductive response to racial disparity.

Finally, even beyond the disparate treatment question, tightening the Guidelines would effectively increase the severity of federal sentencing law, which would disparately *impact* Black men because of their disproportionate involvement in the justice system. By allowing more downward departures for both Blacks and Whites, *Booker* has somewhat softened the Guidelines' rigidity. If

Congress is concerned about the extraordinary rates of incarceration of Black men, increasing severity across the board is no solution.

Notes

- * The author thanks Nick Bagley, Kristina Daugirdas, Paul Hofer, Julian Mortenson, and Marit Rehavi for their helpful comments.
- ¹ U.S. Sentencing Comm'n, Report on the Continuing Impact of *United States v. Booker* on Federal Sentencing (2013), pt. A, at 108. See *Booker*, 543 U.S. 220 (2005); *Kimbrough*, 552 U.S. 85 (2007); *Gall*, 552 U.S. 38 (2007).
- ² U.S. Sentencing Comm'n. *supra* note 1, at 75, 98.
- ³ Sonja B. Starr & M. Marit Rehavi, *Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of Booker*, 123 Yale L.J. ____ (forthcoming fall 2013).
- ⁴ U.S. Sentencing Comm'n, *supra* note 1, pt. E, at 7.
- ⁵ For a more detailed analysis of this point, see Starr & Rehavi, *supra* note 3, pt. II.
- ⁶ M. Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences* (U. Mich. Law & Econ., Working Paper No. 12-002), available at <http://SSRN.com/Abstract=1985377>. Our main analysis focused on non-drug cases, a point the Commission critiques, see U.S. Sentencing Commission, *supra* note 1, pt. E, at 13 n.25, but we included additional results showing that mandatory minimums appeared equally important in shaping disparities in drug cases.
- ⁷ E.g., Marc L. Miller, *Domination & Dissatisfaction: Prosecutors as Sentencers*, 56 Stan. L. Rev. 1211, 1252 (2004).
- ⁸ For instance, suppose Black defendants received departures ten times as often as White defendants, but when they did receive them, they were only half as large. In this (extreme) hypothetical, Black defendants average five times as large a departure as Whites do, but the Commission's analysis would say they were worse off because it would ignore the frequency difference. The Commission does separately analyze departure status as an outcome variable, which is helpful, but including

this analysis does not solve the problem with their main estimates of sentence-length disparity.

- ⁹ See Starr & Rehavi, *supra* note 3, pt. II (broadly critiquing the existing scholarship).
- ¹⁰ This need not be the only objective; one could also focus on structural features of criminal law that have disparate impacts on Black defendants. But identifying the (apparent) "disparate treatment" component can still usefully inform potential policy responses, even if it is not the only concern.
- ¹¹ E.g., Kate Stith, *The Arc of the Pendulum: Judges, Prosecutors, and the Exercise of Discretion*, 117 Yale L.J. 1420, 1430 (2008); Albert W. Alschuler, *Disparity: The Normative and Empirical Failure of the Federal Guidelines*, 58 Stan. L. Rev. 85, 117 (2005).
- ¹² U.S. Sentencing Comm'n, *supra* note 1, pt. A, at 7-8.
- ¹³ The causal claim that *Kimbrough* and *Gall* increased racial disparity is particularly counterintuitive, given that the most direct effect of *Kimbrough* was to benefit (disproportionately Black) crack defendants.
- ¹⁴ Starr & Rehavi, *supra* note 3.
- ¹⁵ We use the time period covered by the Commission's 2010 report, which estimated an even more dramatic disparity for the last period (23%). U.S. Sentencing Comm'n, *Demographic Differences in Federal Sentencing Practices*, 22, fig. C (2010).
- ¹⁶ Even though our measure does not capture arrest disparities, RDD can produce valid estimates of changes in disparity so long as any arrest disparities did not change suddenly just after *Booker*.
- ¹⁷ Starr & Rehavi, *supra* note 3, tbl. 2.
- ¹⁸ *Id.*, fig.1.
- ¹⁹ *Id.*, tbl.2 & fig. 4c.
- ²⁰ *Id.*, tbl.2 & figs. 2a, 2b, 3a, 3b, 3c, 4a, 4b, & 4c.
- ²¹ See also Joshua B. Fischman & Max Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. Emp. Legal Stud. 730 (2012) (using different methods but also reporting results that contradict the Commission).
- ²² 551 U.S. 338 (2007).