

EDITORIAL INTRODUCTION

RACIAL DISPARITY IN WAKE OF THE BOOKER/FANFAN DECISION

Racial disparity under the federal sentencing guidelines pre- and post-*Booker*

Lessons not learned from research on the death penalty

Raymond Paternoster

University of Maryland

The article by Ulmer, Light, and Kramer (2011, this issue) and the corresponding policy essays by Albonetti (2011, this issue), Engen (2011, this issue), Scott (2011, this issue), and Spohn (2011, this issue) in this section of *Criminology & Public Policy* examine the effect of several U.S. Supreme Court decisions on sentencing disparity under the federal sentencing guidelines. In 1984, Congress enacted the Sentencing Reform Act, which created the United States Sentencing Commission (USSC). One motivation for the Act was the belief that too much discretion was provided to judges in the federal system and that as a result there was great disparity in sentencing White and minority defendants.¹ The USSC was given the task of developing and implementing sentencing guidelines for federal judges as a means of controlling judicial discretion, with the goal of achieving greater “uniformity” in sentencing. Prior to the guidelines, federal judges had virtually unlimited discretion to impose sentences so long as they met broad statutory requirements. Under the guidelines, however, the judge had to calculate a defendant’s criminal history and offense level score under strict rules, the result of which was the placement of the defendant on a sentencing grid. The sentence found in the grid was the presumptive sentence, and although departures could be made, the reason for the departure had to be given either in open court or in a written judicial opinion. Furthermore, to monitor and ensure compliance with the guidelines, the Reform Act also provided for appellate review of any departures from the

I am very grateful to Shawn Bushway, Tom Loughran, and Bobby Brame for reading a draft or two of this and for providing helpful conversation. Direct correspondence to Raymond Paternoster, Department of Criminology & Criminal Justice, University of Maryland, College Park, MD 20742–8235 (e-mail: rpaternoster@crim.umd.edu).

1. Other, more conservative, critics also argued that there was too much discretion given to federal judges, but the unwanted product it produced was leniency or “softness” in sentences rather than disparity.

guidelines, with sentences imposed within the calculated guideline range reviewable only under a “clear error” standard (a *de novo* standard of appellate review). In sum, recognizing that “unwarranted disparity caused by broad judicial discretion is the ill that the Sentencing Reform Act seeks to cure,”² Congress attempted to reduce discretion and disparity by trying to make federal sentencing practices comply with strict legal formula. Mandatory guidelines and an appellate review that strictly scrutinizes for compliance with those standards seem like a pretty good way to reduce disparity in sentencing, a suspicion supported by research that includes previous empirical work by the authors of the policy essays (Albonetti, 2011; Engen, 2011; Scott, 2011; Spohn, 2011) in this section (Bushway and Forst, 2011; Bushway and Piehl, 2001, 2007; Engen and Gainey, 2000; Frase, 2005; Reitz, 2005; Spohn, 2000; Tonry, 1996; Ulmer, 2000).^{3,4}

How do we know whether mandatory guidelines were successful in reducing sentencing disparity in federal courts? According to a 2010 study conducted by the USSC, a study that serves as the backdrop for the article by Ulmer et al. (2011) in this section, when the mandatory nature of the guidelines was removed and made advisory by the Supreme Court in *United States v. Booker* (2005) and when the review standards were relaxed and made more deferential by *Rita v. United States* (2007) and *Gall v. United States* (2007), sentencing disparity dramatically increased. This finding was not surprising to me for two reasons.

First, the guidelines that were created sharply reduced judicial discretion at sentencing by providing a reasonably narrow range within which the judge’s sentence was supposed to fit, and any departure from the presumptive sentence was subject to strict appellate review. It was clear that judges (although importantly not prosecutors) had much less discretion under mandatory guidelines and much more sentencing discretion after *Booker*. We should not be too terribly surprised, then, that when sentencing guidelines that were carefully crafted to reduce judges’ sentencing discretion, that were mandatorily imposed, and that were ultimately subject to a rigorous review are now only advisory and the review more deferential that greater disparity in sentencing would result.

Second, as someone who has done state capital punishment but not federal sentencing research, I was not at all surprised by the USSC’s findings of greater racial disparity in sentencing after *Booker*. What the Court did in *Booker*, *Rita*, and *Gall* was to insist that federal sentences be both uniform, that is, consistent across persons, and reflect the unique

2. Stephen S. Trott, Letter to Hon. William W. Wilkins (April 7, 1987); the letter is reprinted at *Federal Sentencing Reporter*, 8: 196 (1995).

3. In their recent review of this area, Bushway and Piehl (2007: 464–465) noted that, “Studies of four jurisdictions using this type of empirical model have found little evidence of racial disparity on the part of judges in strict guideline systems . . . but evidence of substantial disparity in the voluntary guideline systems These results are largely consistent with the general claim that presumptive sentencing guidelines reduce judicial discretion and racial disparity.”

4. To be completely fair, in her excellent review of the literature on sentencing guidelines Spohn (2000) is a little more ambiguous as to whether or not they reduce disparity.

culpability of individual offenders. Although sentencing guidelines were supposed to create consistency and greater uniformity in federal sentences, *Gall* admonished that reviewing courts cannot simply assume that the guideline ranges are reasonable but must “make an *individualized assessment* based on the facts presented” (pp. 596–597). The problem, of course, is that sentences can become more uniform and consistent only at the expense of individual uniqueness and that sentences can be tailored to an individual’s unique culpability only by reducing consistency. The requirement that both principles must be honored is precisely what the Supreme Court did with absolutely no success in *Gregg v. Georgia* (1976) and its companion cases. In *Gregg* (and *Proffitt v. Florida* [1976] and *Jurek v. Texas* [1976]), the Court approved guided discretion statutes because the enumeration of statutory aggravating and mitigating circumstances together with appellate review promised to make death sentences more even handed and less capricious. At the same time, they pointedly struck down mandatory death sentencing schemes because they treated persons “not as uniquely individual human beings, but as members of a faceless, undifferentiated mass” (*Woodson v. North Carolina* [1976: 281]). As a result, death sentences had to be both consistent and individualized. The consequence of trying to conform to both consistency and individualized sentencing was that death sentences under post-*Gregg* procedural reforms were as disparate as those under the denounced standardless system condemned in *Furman v. Georgia* (1972). The comparable Scylla and Charybdis that federal judges have to maneuver through post-*Booker* is the need to adhere simultaneously to the principle of consistency in sentencing in order to ensure uniformity in sentences and the need to tailor the sentence to the unique culpability of individual offenders in the review of such sentences. The USSC’s 2010 report simply confirms what capital punishment researchers had found, that it is difficult in practice to reconcile these two principles and that increased sentencing disparity is the likely result.

The USSC’s 2010 report showing greater racial disparity in federal sentences after *Booker* and *Gall* seems to me, then, to be both believable and predictable. In their article that anchors this issue, Ulmer et al. (2011) offer “an alternative analysis to the USSC’s 2010 report” based on “different analytical and modeling choices.” In their article, they show that racial disparity is more prominent in the prison/probation decision than in the sentence length decision, greater among immigration offenses than among nonimmigration crimes, and greater when the post-*Booker* period is compared with the sentencing regime covered under the PROTECT Act (2003) than the immediate pre-*Booker* period. As the results of the two studies, by USSC and Ulmer et al., seem to conflict with one another, with the perhaps impression that one must be “right” and the other “wrong,” they bear brief discussion.

First, what the USSC (2010) did in its analysis, although involving different “methodological choices” than Ulmer et al. (2011), seems eminently reasonable and their results are due a great deal of respect. Consistent with sound regression discontinuity designs, the USSC analysis compared post-*Booker* racial disparity with that which existed in the

most proximate time period before *Booker*. Furthermore, in modeling the prison/probation decision along with the length of sentence decision, the USSC was assuming that the two decisions were related to one another and to that extent they deserved to be treated jointly. Ulmer et al.'s different methodological choices included a decision to compare different prior time periods with the post-*Booker* period than that used by the USSC. They find that the amount of disparity post-*Booker* depends on which prior period it is compared with. This is interesting, although not surprising, and although it may put bounds on what might be happening in the postguideline period, it does not invalidate the USSC finding that disparity has increased relative to the most recent period. Ulmer et al. also made the choice of separately analyzing subsets of the population that the USSC had, excluding immigration cases from one of their analyses. What they found was greater disparity among immigration than among nonimmigration cases, confirming the frequent finding that a general "treatment" effect may not hold for all strata within a population. It seems to me that what they have done here, however, is to elaborate and extend the USSC's findings rather than raise a suspicion about them.

Ulmer et al. (2011) also preferred to conduct separate analyses of the prison/probation and the length of sentence decision, finding greater disparity in the former than in the latter. Without making any judgment about the USSC's (2010) choice to analyze both in the same equation, it is worth thinking about the implications of Ulmer et al.'s strategy to conduct a separate analysis. Ulmer et al. are asserting that judges first make a decision whether to sentence a defendant to prison or probation, and then completely independent of this decision, they decide how much time the person should be sentenced for. This assertion seems to me to be a very strong claim about how sentencing is actually done by judges. Whatever disagreements one may have with the USSC's decision to treat the two decisions, the model of sentencing presumed by Ulmer et al. should be completely understood.

In sum, what does an outsider make of the USSC's (2010) findings and those of Ulmer et al. (2011)? Because of my experience with the U.S. Supreme Court's attempt to regulate state death sentences, I was not surprised by the fact that the USSC found greater disparity after the *Booker* and *Gall* decisions. Although I now have a more nuanced understanding⁵ of this thanks to Ulmer et al.'s extensive analyses, nothing in their report would lead me to change my opinion that a consequence of moving from mandatory to advisory federal sentencing guidelines with a more relaxed standard of review is greater racial disparity in sentencing. What both sets of studies would seem to agree on is that when the federal sentencing regime became advisory, disparity in sentencing increased, with a disagreement mainly about the magnitude of that increase. But this is not the whole story.

5. Spohn (2011) concluded that what we have here with Ulmer et al. and the USSC's report are "two major studies [that] have reached different conclusions." I do not necessarily see it quite that way. I see two analytically/methodologically very different studies that have come to slightly different conclusions that complement one another.

Many practitioners and scholars did not like the mandatory federal sentencing standards because they were too harsh and retributive, and they may fear that empirical findings of greater disparity under an advisory regime such as those reported by the USSC may hasten their return. That battle will have to be fought on its own ground, however, and not on the grounds that advisory standards will not result in greater sentencing disparity.

So then, “Why deny the obvious child?” Although I was not surprised that the USSC’s 2010 report found an increase in disparity in federal sentences after *Booker* and *Gall*, I was surprised by the denial that permeates the policy essays that join Ulmer et al.’s (2011) article. I have noted that these policy essay authors (Albonetti, 2011; Engen, 2011; Scott, 2011; Spohn, 2011) are part of a sentencing literature in criminology that has concluded that sentencing guidelines, when they are mandatory and strictly monitored, have reduced disparity. Based on their own previous excellent work on guidelines and the observations they presented in their essays, it is not clear why these authors would be surprised by, or why they would deny, the USSC’s findings. It is difficult to tell whether this reaction is based on genuine doubt that *Booker* and *Gall* resulted in greater disparity, or more on the aforementioned fear that these findings might be used to bring back some variation of the old mandatory/strict review regime.

For example, both Albonetti’s (2011) and Spohn’s (2011) policy essays provide excellent historical context to the creation of the U.S. Sentencing Commission, and like any good history, they help us understand the present. They show that there was great dissatisfaction with what was perceived to be an unduly large amount of disparity in federal sentences, and consensus that this disparity resulted from two features of the existing sentencing regime: (a) judges had extensive sentencing discretion, and (b) sentences were not generally reviewable. From another direction, however, there was equal angst at the time that federal sentences, particularly with respect to drug and gun crimes, were too lenient. As Albonetti correctly points out, the 1984 Sentencing Reform Act was not the only sentencing-related piece of legislation passed by Congress at that time; there was also the Anti-Drug Abuse Act of 1986, the Comprehensive Crime Control Act of 1984, and the Omnibus Anti-Drug Act of 1988, all of which were expressions of a desire to get tougher on crime.

Both the desire to minimize sentencing disparity and the desire to punish more certainly and harshly could be satisfied by a sentencing regime that provided judges with fairly narrow sentencing ranges that increased when the offense was more serious and the offender’s criminal history was more extensive, and which required appellate review when there were departures from this—ta da . . . mandatory sentencing guidelines with limited ability for a judge to go outside those guidelines. One agreed upon consequence of the 1984 Sentencing Reform Act is exactly what Albonetti (2011) noted in her essay, it “virtually transformed sentencing practices. Policy priorities were aimed at severely limiting judicial discretion in an attempt to eliminate unwarranted sentence disparity.” In her essay, Spohn (2011) made precisely the same observation, that the Sentencing Reform Act was born from both liberal and conservative political sentiments and that the mandatory regime with appellate

review devised and implemented by the USSC was intended to curtail disparity sharply. She says, “[i]t is also clear that the guidelines were designed to eliminate discrimination based on legally irrelevant characteristics of the offender. . . . the potential for racial and ethnic discrimination was limited by the fact that the guidelines were mandatory and that judge-initiated departures were regulated closely.” The first point I want to make about this is that because the mandatory guidelines were created to “curtail disparity sharply” and that they “virtually transformed sentencing practices,” why should anyone be surprised that when they were both made advisory and less stringently scrutinized after *Booker* and *Gall* that greater disparity would result? The second point is that as liberals, who wanted to get rid of disparity, entered into a Faustian bargain with conservatives, who wanted to get tougher on crime, why would anyone be surprised by the fact that mandatory guidelines based on offense seriousness and offender criminal history scores tended to be punitive?

It would seem to an outsider like me that given the USSC’s (2010) findings about greater disparity immediately after *Booker*, and the fact that all parties seem to agree that the least amount of sentencing disparity was found during the PROTECT regime when sentencing guidelines were both mandatory and stringently reviewed, that if one really wanted to minimize discretion, they would want to return to that kind of regime. However, as we will see in a moment, *no one* among the current set of authors wants to do that, primarily it seems because they do not wish to return to a regime that they believe is unnecessarily harsh and severe. Before getting to that issue, however, a few more observations need to be made.

I found myself in agreement with certain of Engen’s (2011) conclusions. His policy essay is very much a painstaking dissection of the different analytical and methodological choices that the authors of the two studies made (Ulmer et al., 2011; USSC, 2010), and he concluded that “certain methodological choices probably affected the conclusions” reached by both studies. I agree. In addition, after reading Engen’s comments, I was convinced that there was likely *more* disparity in the post-*Booker* period than what Ulmer et al. reported. He raised the issue that the decision of Ulmer et al. to include in their model both the defendant’s criminal history and the presumptive sentence might be suppressing a larger race effect, a possibility that is often discussed in this literature and that even Ulmer et al. acknowledged: “Furthermore, the criminal records of Black males may themselves be the product of discriminatory processes.”

An additional conclusion reached by Engen (2011) that bears repeating is that “[t]he general unavailability of data on charging and plea-bargaining remains . . . *the greatest challenge to the validity of sentencing research*” (emphasis added). As all parties to this discussion acknowledge, what is lacking from both the USSC’s (2010) and Ulmer et al.’s (2011) analysis is information about prosecutors’ decisions. Another lesson not learned from capital punishment research was that the lion’s share of the disparity in how criminal cases are handled often occurs long before sentencing takes place—in the hands of the prosecutor. Countless empirical studies of how states impose the death penalty have shown that there is substantial disparity at the decision of the prosecutor to charge a crime as capital, and there

is minimal disparity at sentencing. State capital punishment research has shown, therefore, that most of the “action” is at the level of the charging decision. Although there seems to be general agreement about the need to consider prosecutorial decision making in federal sentencing research (see Bushway and Piehl, 2007), there seems to be no urgency in doing something about it. There is instead great lamentation about the lack of this data and its implications, and an internecine quibble over the results of analyses of conviction data sets. I would suggest that because both parties to this issue (USSC and Ulmer et al.) have only conviction data, the argument here is about a small part of the story of disparity in federal sentencing either with or without mandatory sentencing guidelines.

Furthermore, it is not as if either of the two sets of authors can blithely dismiss this omission of prosecutorial information by making the claim that they are only speaking to disparity in *sentencing* as (again, Engen [2011] nicely points this out) discrimination at earlier points in the system distorts the types of cases that are considered at sentencing. It seems a bit odd to have such sharp arguments and disagreements over the results of sentencing studies that fail to consider prosecutorial charging decisions. Hopefully, at some point, the hand wringing and the “weeping, wailing, and gnashing of teeth” will stop and something will be done to collect good prosecutorial decision-making data at the federal level.

Scott’s (2011) policy essay is somewhat of the dissident in this series of essays because he chose to focus mainly on the common ground between the USSC (2010) report and the Ulmer et al. (2011) article rather than on the differences, but I think a reminder of the great deal of common ground is very well served: “in several of their key findings, the Commission’s research and the new analysis by Ulmer et al. (2011) reach similar results.” The common ground identified by Scott is that in both studies: (a) the imprisonment decision disadvantages Black males in comparison with White males, (b) evidence of racial disparity under the mandatory guidelines before 2003 was fragile, and (c) race disparity against Black males was at its lowest level “when the Guidelines were at their most mandatory and inflexible” and when there was strict appellate review, under the PROTECT Act regime. This last point is very interesting because how to respond to it was something that all four policy essays had in common.

As I suggested a few paragraphs earlier, if one is interested in driving sentencing disparity down to its minimal level, the evidence would seem to indicate that you have mandatory sentencing guidelines and a very stringent appellate review—the regime under the PROTECT Act. It makes complete sense. However, *none* of the authors of these essays settles on that conclusion. Albonetti (2011), for example, after acknowledging that mandatory guidelines and strict review, “virtually transformed sentencing practices” nevertheless refuses to believe that making them advisory and the review more deferential would result in substantially more disparity and that “[t]here is no need to institute statutory remedies for sentences that do not greatly differ from those imposed under pre-*Booker* mandatory guidelines structure.” Essentially, she dismisses the USSC (2011) and

accepts Ulmer et al. (2011), thinking that judges essentially behave the same pre and post-*Booker*, even though she acknowledged that mandatory guidelines “transformed sentencing practices.” Despite making virtually a similar argument about the power of mandatory guidelines and strict review in sharply reducing disparity, Spohn (2011) too does not argue for their return. We start to get a sense of what the real opposition is in Engen’s (2011) essay where he refers to the fact that “many observers, including federal judges, believe [the guidelines] are unjust.” He goes on to wonder what should be made of strict consistency or disparity in sentencing under “laws [that] are indeed unfair.” In the very next paragraph, Oz’s curtain is lifted when he implies that greater disparity may not be such a bad thing if “with increased discretion, judges will hand down sentences that most observers agree are more ‘appropriate,’ on average, than if they had followed the guidelines closely, and that all races, ethnicities, and genders will benefit from this discretion.” More appropriate in what sense? Because sentences are less severe when guidelines are advisory? Is the real problem with mandatory guidelines and stringent review, then, not disparity but substantive injustice as Engen argues? If it is, then (a) that is not an empirical question that can be addressed by scientific studies, and (b) perhaps we should bring it out into the open and address it directly. Scott (2011) also saw a return to the regime under the PROTECT Act as a logical policy alternative, noting that it was the policy choice recommended by Justice Stevens in his *Booker* dissent, and Scott provides sound reasons as to why this route makes sense if you want to diminish disparity. Like the others, however, Scott dismisses this policy remedy for myriad reasons.

An outsider to this debate about lesser or greater disparity after the *Booker* decision gets the sense that at least to some degree we are not really talking about how much more or less disparity exists under mandatory versus advisory sentencing guidelines. The uniformity and resoluteness with which a return to mandatory guidelines and strict review are rejected by the policy essay authors in this section, and the only slightly veiled appreciation for greater substantive justice under advisory guidelines that permits greater discretion to judges, both lead me to suspect that at least part of the issue among federal sentencing scholars is that the mandatory guidelines were too severe for their taste and they fear any attempt to return to them. Research findings such as those in the USSC 2010 report, it is suspected, would naturally lead in that direction. In my language, given a choice between the greater consistency of *Gregg* or the greater discretion and ability to individualize sentences of *Woodson*, they definitely prefer the latter. That, however, seems a preference that empirical research cannot address.

References

- Albonetti, Celesta A. 2011. Judicial discretion in federal sentencing: An intersection of policy priorities and law. *Criminology & Public Policy*. This issue.
- Bushway, Shawn and Brian Forst. 2011. Studying discretion in sentencing. Working paper.

- Bushway, Shawn D. and Anne Morrison Piehl. 2001. Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, 35: 733–764.
- Bushway, Shawn D. and Anne Morrison Piehl. 2007. Social science research and the legal threat to presumptive sentencing guidelines. *Criminology & Public Policy*, 6: 461–482.
- Engen, Rodney L. 2011. Racial disparity in the wake of *Booker/Fanfan*: Making sense of “messy” results. *Criminology & Public Policy*. This issue.
- Engen, Rodney L. and Randy Gainey. 2000. Conceptualizing the role of legal and extra-legal factors under sentencing guidelines: Reply to Ulmer. *Criminology*, 38: 1245–1252.
- Frase, Richard S. 2005. Sentencing guidelines in Minnesota, 1978–2003. In (Michael H. Tonry, ed.,) *Crime and Justice: A Review of Research*, vol. 32. Chicago, IL: University of Chicago Press.
- Reitz, Kevin R. 2005. The enforceability of sentencing guidelines. *Stanford Law Review*, 58: 155–174.
- Scott, Ryan W. 2011. Race disparity under advisory guidelines: Dueling assessments and potential responses. *Criminology & Public Policy*. This issue.
- Spohn, Cassia. 2000. Thirty years of sentencing reform: The quest for a racially neutral sentencing process. In *Policies, Processes and Decisions of the Criminal Justice System*. Vol. 3, *Criminal Justice 2000*. Washington, DC: U.S. Department of Justice.
- Spohn, Cassia. 2011. Unwarranted disparity in the wake of the *Booker/Fanfan* decision: Implications for research and policy. *Criminology & Public Policy*. This issue.
- Tonry, Michael H. 1996. *Sentencing Matters*. New York: Oxford University Press.
- Ulmer, Jeffery T. 2000. The rules have changed—so proceed with caution: A comment on Engen and Gainey’s method for modeling sentencing outcomes under guidelines. *Criminology*, 38: 1231–1243.
- Ulmer, Jeffery T., Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC’s 2010 report. *Criminology & Public Policy*. This issue.
- U.S. Sentencing Commission. 2010. Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report’s Multivariate Regression Analysis. Washington, DC: Author.

Court Cases Cited

- Furman v. Georgia*, 408 U.S. 238 (1972).
- Gall v. United States*, 552 U.S. 38 (2007).
- Gregg v. Georgia*, 428 U.S. 153 (1976).
- Jurek v. Texas*, 428 U.S. 262 (1976).
- Proffitt v. Florida*, 428 U.S. 242 (1976).
- Rita v. United States*, 551 U.S. 338 (2007).
- United States v. Booker*, 543 U.S. 220 (2005).
- Woodson v. North Carolina*, 428 U.S. 280 (1976).

Statutes Cited

Anti-Drug Abuse Act of 1986, Pub.L. No. 99-570, § 1105(a), 100 Stat.3207-11 (1986).
Comprehensive Crime Control Act of 1984, S. Rep. No. 98-225, at 1312 (1983); H.R. Rep. No. 98-1017, at 55-56 (1984).
Omnibus Anti-Drug Act of 1988, Pub.L. No. 100-690, § 6371, 102 Stat. 4181, 4370 (1988).
PROTECT Act of 2003, Pub.L. 108-21, 117 Stat. 650, S. 151, enacted April 30, 2003; amendment Stat. 667.
Sentencing Reform Act of 1984, 18 U.S.C. §§ 3551-3626 and 28 U.S.C. §§ 991-998, as amended 1985-1988, 1990, 1992, 1994, and 1996.

Raymond Paternoster is professor in the Department of Criminology & Criminal Justice at the University of Maryland. His research interests include offender decision making, capital punishment, and the application of quantitative methods to criminological research.