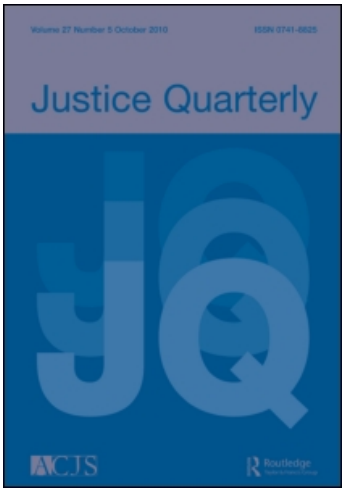
This article was downloaded by: *[Ulmer, Jeffery T.][Pennsylvania State University]*
On: *7 March 2011*
Access details: *Access Details: [subscription number 933509531]*
Publisher *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Justice Quarterly
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722354

The "Liberation" of Federal Judges' Discretion in the Wake of the Booker/Fanfan Decision: Is There Increased Disparity and Divergence between Courts?

Jeffery Ulmer; Michael T. Light; John Kramer

First published on: 07 March 2011

To cite this Article Ulmer, Jeffery , Light, Michael T. and Kramer, John(2011) 'The "Liberation" of Federal Judges' Discretion in the Wake of the Booker/Fanfan Decision: Is There Increased Disparity and Divergence between Courts?', Justice Quarterly,, First published on: 07 March 2011 (iFirst)
To link to this Article: DOI: 10.1080/07418825.2011.553726
URL: http://dx.doi.org/10.1080/07418825.2011.553726

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

JUSTICE QUARTERLY 2011, 1–39, iFirst Article

The "Liberation" of Federal Judges' Discretion in the Wake of the Booker/Fanfan Decision: Is There Increased Disparity and Divergence between Courts?

Jeffery Ulmer, Michael T. Light and John Kramer

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

The US Sentencing Guidelines are among the most ambitious attempts in history to control sentencing discretion. However, a major sea change occurred in January of 2005, when the US Supreme Court ruled in *United States v. Booker* and *Fanfan*, that in order to be constitutional, the federal guidelines must be advisory rather than presumptive. The impact of the Booker/Fanfan decisions on interjurisdictional variation and sentencing disparity is an opportunity to examine the issue of whether the increased opportunity to sentence according to

Jeffery T. Ulmer is currently Associate Professor of Sociology and Crime, Law, and Justice. His articles in journals such as *American Sociological Review*, *Social Problems*, *Criminology*, *Criminology and Public Policy*, *Justice Quarterly*, *The Sociological Quarterly* and *Journal of Research in Crime and Delinquency* on courts and sentencing, criminological theory, race, ethnicity, and violence, religion and crime, symbolic interactionism, criminal enterprise, and the integration of qualitative and quantitative methods. He is the author of *Social Worlds of Sentencing: Court Communities Under Sentencing Guidelines* (1997, State University of New York Press), and co-author (with Darrell Steffensmeier) of *Confessions of a Dying Thief: Understanding Criminal Careers and Illegal Enterprise* (2005, Aldine-Transaction) which won the 2006 Hindelang Award from the American Society of Criminology. His newest book (with John Kramer), *Sentencing Guidelines: Lessons from Pennsylania* was published in 2009 by Lynne Rienner Publishers. Michael T. Light is a PhD candidate in the Department of Sociology & Crime, Law and Justice at The Pennsylvania State University. His major research interests include punishment, immigration, race, ethnicity and criminal offending, and spatial analysis of crime. He is currently studying the impact of citizenship status on sentencing outcomes and the spatial processes of White and Black violent crime. John H. Kramer is Professor of Sociology and Crime, Law, and Justice. He served as Executive Director of the Pennsylvania Commission on Sentencing from 1978 to 1998, Advisor to the U.S. Sentencing Commission from 1995 to 1996, and Staff Director of the U.S. Sentencing Commission from 1996 to 1998. His research and policy development focus have been in the areas of sentencing disparity, the impact of sentencing reform, and evaluating the effectiveness of intermediate punishment and drug courts. His research has been published in numerous journals including *Criminology, Justice Quarterly, Social Problems*, *Journal of Research in Crime and Delinquency,* and *Criminology and Public Policy.* He is co-author (with Jeffery Ulmer) of *Sentencing Guidelines: Lessons from Pennsylvania* (2009, Lynne Rienner Publishers). Correspondence to: Jeffery Ulmer, Department of Sociology/Crime, Law, and Justice, Penn State University, 211 Oswald Tower, University Park, PA 16802, USA. E-mail: jtu100@psu.edu

ISSN 0741-8825 print/1745-9109 online/11/000001-39
© 2011 Academy of Criminal Justice Sciences
DOI: 10.1080/07418825.2011.553726

Routledge
Taylor & Francis Group

substantively rational criteria entails increased extralegal disparity. We draw on a conceptualization of courts as communities and a focal concerns model of sentencing decisions to frame expectations about federal sentencing in the wake of Booker/Fanfan. We test these expectations using USSC data on federal sentencing outcomes from four time periods: prior to the 2003 PROTECT Act, the period governed by the PROTECT Act, post-Booker/Fanfan, and post-Gall v US. In general, we find that extralegal disparity and between-district variation in the effects of extralegal factors on sentencing have *not* increased post-Booker and Gall. We conclude that allowing judges greater freedom to exercise substantive rationality does not necessarily result in increased extralegal disparity.

Keywords sentencing; judicial discretion; federal courts; U.S. v. Booker decision; U.S. v. Gall decision

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

The US Sentencing Guidelines, put into effect 1 November 1987, explicitly strove to eliminate "unwarranted disparity" in sentences based on race, gender, ethnicity (i.e., extralegal factors) as well to create more uniformity across federal court jurisdictions. The architects of the guidelines specifically "sought uniformity in sentencing by narrowing the wide disparity in sentences imposed by different federal courts for similar criminal conduct by similar offenders" (USSG, 2004: §1A1.1, p.s.: 2).

United States v. Booker (543 U.S. 220 {2005} and a joined case *United States v. Fanfan* 542 U.S. 296 {2004}, hereafter *Booker*) ruled that the Federal Sentencing Guidelines (hereafter, guidelines) would henceforth be advisory rather than presumptive in federal sentencing. Subsequent decisions, especially *Gall v. United States* (128 S. Ct. 586 {2007}), have also expanded federal judges' discretion. Many fear these decisions have brought increased "entropy," that is less consistency and uniformity, in sentencing between district courts, and also increased unwarranted disparity based on the social status characteristics of defendants (see reviews by Frase 2007; Hofer, 2007). Hofer (2007) argues that if a primary goal of federal sentencing reform was a reduction of unwarranted disparity, the impact of Booker on disparity is among the most important questions facing sentencing policy-makers. However, a plausible case can be made that despite the fact that the Guidelines are now advisory, federal sentencing disparity may not change dramatically. Most judges who served pre-Guidelines are no longer active, and their replacements may have become accustomed to relying on the Guidelines as a benchmark (Kramer, 2009).

Examining the impact of Booker on federal sentencing disparity also is an opportunity to address important theoretical issues. Sentencing practices have long exhibited the tension between formally rational reform efforts designed to promote uniformity and control discretion and substantively rational interests in flexibility, individualized (and localized) sentencing (Savelsberg, 1992; Walker, 1993). However, the discretion to individualize sentences according to substantively rational criteria also allows the potential for wide variation between jurisdictions in sentencing severity and decision criteria as well as disparity

between defendants on the basis of social statuses (Kramer & Ulmer, 2009). The risk of unwarranted disparity is said to be the cost of allowing court actors the discretion to individualize, and localize, sentencing and to consider value-based or ideological sentencing criteria that are extraneous to sentencing guidelines. The impact of *Booker* and its aftermath is an opportunity to examine whether increased opportunity to sentence according to substantively rational criteria entails increased disparity and increased between-court variation.

Engen (2009) noted the paucity of research on what happens in the wake of the repeal or relaxation of presumptive sentencing schemes and called for research that examines how the exercise of discretion is related to the structure of (and changes in) sentencing laws and guidelines. By examining sentencing in the aftermath of *Booker* and *Gall*, which dramatically loosened constraints on federal judicial discretion, we are helping to address this issue. Using federal sentencing data, we investigate:(1) whether variation in sentencing has increased between district courts since *Booker* and *Gall*, (2) whether extralegal disparity has increased on the basis of race/ethnicity and gender, and (3) whether inter-district variation in the effects of extralegal characteristics has increased. In doing so, we also account for the effects of several district court characteristics that have been shown in pre-*Booker* research to be important sentencing influences.

Research on Federal Sentencing Disparity

Research on punishment outcomes has been a major enterprise in criminology and sociology, and there is a long tradition of research investigating sentencing disparities along racial, ethnic, and gender lines. Reviews of state and federal sentencing research show legally prescribed factors are the strongest predictors of sentencing outcomes, but defendant social statuses often influence the likelihood and type of incarceration, sentence length, and guideline departures, and that court contextual factors both directly influence sentencing and condition the effects of individual-level factors (e.g., Hartley, Maddan, & Spohn, 2007; Johnson, 2005, 2006; Johnson, Ulmer, & Kramer, 2008; Kramer & Ulmer, 2009; Mitchell & MacKenzie, 2004; Spohn, 2000; Zatz, 2000).

Much scholarly research on federal courts has assessed disparity among individual offenders under the Guidelines prior to *Booker* (for a review, see USSC, 2004). These studies have often found small-to-moderate racial and ethnic sentencing differences benefitting whites, though Guideline-relevant factors exert much larger effects than offender status characteristics (e.g., Albonetti, 1997; Johnson et al., 2008; Mitchell & MacKenzie, 2004; Mustard, 2001; Steffensmeier & Demuth, 2000; USSC, 2004). Research using pre-*Booker* data has also shown that young minority males are particularly disadvantaged in incarceration and sentence length decisions (Doerner & Demuth, 2010), that defendant race, age, and gender influenced prosecutorial charge reductions which in turn influence sentencing outcomes (Shermer & Johnson, 2010), and

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

that the influence of race/ethnicity and gender on sentencing varies by judge (Anderson & Spohn, 2010).

Pre-*Booker* research has detected considerable between-district variation in sentence severity, practices surrounding guideline departures, and in the effects of extralegal variables on sentencing, suggesting that federal district courts form distinctive court communities (Feldmeyer and Ulmer, in press; Hartley et al., 2007; Johnson et al., 2008; Kautt, 2002, 2009; Ulmer, 2005; Ulmer, Eisenstein, & Johnson, 2010). However, with the exception of two recent USSC reports, all of the published research reporting moderate extralegal federal sentencing disparity is based on pre-*Booker* data and the overwhelming majority on pre-PROTECT Act data. With the increase in judicial discretion post-*Booker*, such extralegal disparities may have increased. Similarly, the significant pre-*Booker* inter-district variation found in sentencing severity and in the effects of extralegal factors on sentencing outcomes may have also increased, confirming the fears of observers such as Richter (2008) that federal courts have become less uniform in their sentencing behavior.

As noted above, what we know about post-*Booker* sentencing comes from two reports published by the USSC in 2006 and 2010. The 2006 report showed that the majority of federal cases continued to be sentenced in conformity with the Guidelines, but that the rate of above-range, government-sponsored below-range, and other below-range sentences increased. Multivariate analyses showed that social status factors were moderately associated with sentence length, but that their effects pre- and post-*Booker* were similar (USSC, 2006). The report also examined conformity and departures by circuit and district from 2001 to January 2006 and concluded that regional sentencing differences have been relatively stable. Likewise, commentary on the USSC report stated: "With a little over a year's experience under *Booker*'s new 'advisory' guidelines regime, the cumulative results can be summarized as 'much ado about nothing, or at least much ado about not very much'" (Thompson, 2006, p. 269). Overall, the 2006 report notes that disparity decreased in the PROTECT era, but reverted back closer to pre-PROTECT act sentencing practices post-*Booker*.

However, the 2010 report (USSC, 2010), which included data up to FY 2009, finds that race disparity has increased in the later post-Booker period, as compared to the PROTECT Act period (the USSC 2010 report did not include the pre-PROTECT years in their comparisons). Their models first replicated the analyses in the 2006 report with the newer data included, and then estimated a "refined model." Their "Booker report model" showed that blacks received about 2% longer sentences than whites (not significant) during the PROTECT period, but that blacks received 7% and 10% longer sentences than whites in the post-Booker and post-Gall periods, respectively. Their "refined model," which did not control for criminal history, but did break out black, white, and Hispanic defendants by gender, found that black males received 5% longer sentences than white males in the PROTECT period, but 21% longer sentences by the late post-Booker period. The 2010 report also found that non-citizens were increasingly sentenced more harshly than the US citizens and that gender disparity fluctuated across time periods.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Our study goes beyond the 2006 and 2010 USSC reports in several important respects. First, the reports focused on the overall effects of legal and extralegal factors pre- and post-Booker. We do this as well, but also focus on the *between-district variation* in the effects of extralegal factors associated with sentencing decisions. As suggested earlier, a key prediction is not just that extralegal disparity will increase generally, but that its patterns will *diverge* more among districts. This question has not yet been directly examined. Thus, we examine the following questions by comparing sentencing outcomes across four time periods: pre-PROTECT Act, post-PROTECT, post-*Booker*, and post-*Gall*:

(1) Has inter-district variation in sentencing outcomes increased?
(2) Has disparity based on defendant social statuses (e.g., race/ethnicity, gender, age, education, and citizenship) increased?
(3) Has inter-district variation in the effects of defendant social status factors on sentencing increased?
(4) The aftermath of Booker/Fanfan: The re-empowering of federal judicial discretion.

From 1996 to 2005, legal developments moderately expanded judicial sentencing discretion, then sharply restricted it, and then, culminating in *Booker*, dramatically expanded it again. From 1987 to 1996, discretion historically resting with the judiciary was tightly constrained, and shifted to the prosecutor (Stith & Cabranes, 1998). Congress continued restricting judges' sentencing discretion during this period, sending directives to the commission, and passing mandatory minimums to be incorporated into the guidelines. Then, in *Koon v. United States* (518 U.S. 81, 99-100 1996), the Supreme Court restored some discretion to judges by establishing an "abuse of discretion" standard for appellate review of departures from the Guidelines. As a result, judicial "unguided" downward departures more than doubled between 1996 and 2001 (Stith, 2008). Congress sought to counter *Koon* with the Feeney Amendment to the PROTECT Act of 2003 (Pub. L No. 108-21, 117 Stat. 650, 667), which replaced the "abuse of discretion" standard for Guideline departures with a "de novo" review and directed the commission to reduce departure mechanisms. In our analyses, we refer to the former period as the "pre-PROTECT Act" and the latter as the "PROTECT" or "post-PROTECT Act" period.

The *Booker* decision in 2005 ruled that mandatory guidelines violated the sixth amendment's guarantee of the right to a jury trial. Judges could not constitutionally assess "real offense" conduct that increased sentences on factors not considered at trial by a jury. The court's solution was that the Guidelines would become advisory only. Judges must consider the Guidelines, but their discretion was returned to at least pre-PROTECT Act, though not to pre-Guideline levels. Stith (2008, p. 1426) states: "In a dramatic exercise of judicial power, *Booker* undid the Feeney Amendment, limited the power that inheres in prosecutors in a regime of mandatory sentencing rules, and counteracted the

centralizing impulse of Main Justice." In our analyses, we refer to this period as the "post-*Booker*" era.

Subsequently, the court enhanced judges' discretion restored in *Booker* by clarifying the meaning of "advisory" in *Rita v. United States* (127 S. Ct. 2456, 2487 2007), where it ruled that federal appellate courts *may* but are not *required* to presume Guideline sentences to be reasonable. Consequently, sentences outside the Guidelines cannot be automatically regarded as unreasonable. In *Gall v. United States* (128 S. Ct. 586 2007), the court rejected appellate case law finding that the greater the departure from the guidelines, the more extraordinary the justification necessary for the departure. What is more, the court in *Gall* held that district judges may not automatically presume the Guideline range to be reasonable and must "make an individualized assessment based on the facts presented" (*Gall v. U.S.*, 128 S. Ct. 586 2207 pp. 596-597, our emphasis). In *Kimbrough v. United States* (128 S. Ct. 558 2007), the court ruled that in crack cocaine cases, judges could reasonably conclude that Guideline sentences were not reasonable in an individual case. Although both defendants and the government maintain the power to appeal federal sentences, in the wake of *Booker* the standard of review now relies on the "reasonableness" of the sentence, and an "abuse of discretion" standard, rather than correct application of the guidelines. Following similar procedures of the USSC, in our analysis, we refer to the 2008–2009 period as the "post-*Gall*" period.[1]

Legal observers have had different reactions to these developments. US Attorney General Eric Holder (2009) noted that the uniformity and the control of judicial discretion, *per se*, do not guarantee justice: "The desire to have an almost mechanical system of sentencing has led us away from individualized, fact-based determinations that I believe, within reason, should be our goal." However, prominent US Attorney John Richter (2008, p. 340), presenting a view held by many other federal prosecutors, argues that "Post-*Booker* sentencing threatens equal justice under law." Richter (2008) fears that the wake of *Booker*, *Gall*, and *Rita* will reduce sentencing consistency between district courts and increase unwarranted disparity. Furthermore, the dissenting opinions of Justices Stevens, Thomas, and Scalia in *Booker* each noted that Congress clearly intended to restrict judicial discretion in order to curb unwarranted disparity in federal sentencing and argued that the court majority's remedy of making the Guidelines advisory would jeopardize that goal.

Does the Return of Substantive Rationality Bring Greater Disparity?

Savelsberg (1992) argued that the Guidelines imposed a regime of strict formal rationality overtop a sentencing process which was traditionally permeated with substantively rational considerations (see also Kramer & Ulmer, 2009). The

1. The post-Gall period includes December 2007 (when the decision was handed down) through September 2009. See "Data" section for further details.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

flexibility inherent in substantive rationality is also said to permit the possibility of bias, discrimination, and unwarranted disparity (Ulmer & Kramer, 1996). It is exactly this flexibility inherent in substantive rationality that has increased post-Booker. After *Booker*, federal prosecutors still retain considerable discretion in the charging and plea agreement processes, and the initiation of substantial assistance departures, though *Booker* and subsequent decisions may have changed the distribution and exercise of local prosecutors' pre-conviction and pre-sentencing discretion in as yet unknown ways. However, the key point for our analysis is that these decisions have indisputably increased judges' sentencing discretion relative to the pre-PROTECT and PROTECT eras (Stith, 2008), and with such discretion, their opportunities to exercise substantive rationality.[2] This opportunity may open the way for increased extralegal disparity, as well as greater inter-district "entropy" in federal sentencing, in which different federal courts sentence similar cases in increasingly different ways.

The focal concerns perspective argues that sentencing decisions are structured by court actors' interpretations of three substantive focal concerns of punishment, each embedded in a local court context: (1) defendant blameworthiness, (2) protection of the community/defendant dangerousness, and (3) practical constraints and consequences (see Kramer & Ulmer, 2009; Steffensmeier & Demuth, 2000; Steffensmeier, Ulmer, & Kramer, 1998). The subjective assessment of focal concerns makes it likely that stereotypes and biases based on race/ethnicity or other extralegal defendant characteristics such as gender, age, education, and citizenship, could influence the sentencing process, *given a context where such stereotypes are prevalent* (Kramer & Ulmer, 2009). If assessments of focal concerns are shaped by attributions based on extralegal statuses, the return of substantive rationality and judicial discretion post-*Booker* would see increased unwarranted disparity.

The influential "liberation hypothesis" of legal discretion is also relevant here (Kalven & Zeisel, 1966; Smith & Damphousse, 1998; Spohn & Cederblom, 1991). This hypothesis argues that extralegal influences on legal decisions increase as the opportunity for discretion increases, and vice versa. In the aftermath of *Booker*, the liberation hypothesis would imply that the influence of extralegal variables on sentencing should become greater post-Booker/Fanfan as well. If judges become "liberated" in their discretion relative to past restrictions, they might exhibit a greater tendency to consider extralegal factors, resulting in greater social status disparities.

2. Savelsberg (1992), among many others, predicted that the Guidelines would not in fact reduce extralegal sentencing disparity, but would "hydraulically" displace it to earlier stages in criminal processing. Recent commentators have also used disparity in pre-conviction prosecutorial practices to question whether presumptive guidelines systems have been as successful as claimed (Bushway and Piehl 2007). On the other hand, Hofer (2007) argues that the advisory guidelines system created by *Booker*, by shifting some discretion back to judges, may have reduced disparity by permitting judges to adjust for differences in prosecutorial practices. Furthermore, Stith (2008) argues that regardless of how prosecutorial discretion might have changed post-Booker, *Booker* has re-empowered judicial discretion, and may have enabled judges to check prosecutorial sentencing influence (Stith, 2008).

Court communities are said to exhibit different, locally varying substantive legal rationalities and interpretations of focal concerns that influence sentencing outcomes and processes at least as much as formal policies and legal structures (Engen, Gainey, Crutchfield, & Weis,. 2003; Kramer & Ulmer, 2009; Savelsberg, 1992). Extralegal disparity is therefore said to be localized and context-dependent (Johnson et al., 2008; Kautt, 2009; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). The relaxation of Guideline constraints post *Booker* therefore might exacerbate variation in sentencing between districts, resulting in greater "entropy" in federal sentencing, as observers such as Richter (2008) fear. As Stith (2008, p. 1427) puts it: "Once again, sentencing is to a significant extent a 'local' event. After *Booker*, the Sentencing Commission and Main Justice may still be calling signals but the decision makers on the playing field—judges and prosecutors—need not follow them."

However, plausible alternative views about the impact of *Booker* exist. For example, Reitz (2005) predicted that the Guidelines would continue to strongly govern federal sentencing in the aftermath of *Booker* and noted that the decision did not reinstitute the level of judicial sentencing discretion that existed in the time before the Guidelines. Further, inter-district variation in federal sentencing existed prior to *Booker* and might change very little as a result of *Booker*. In a similar vein, Wooldredge's (2009) analysis of long-term sentencing trends under Ohio's sentencing guidelines found general stability in sentencing patterns over time with respect to racial and gender disparity. Based on his findings, Wooldredge (2009) speculated that Ohio's recent shift to voluntary (as opposed to presumptive) guidelines would likely produce few substantive changes in disparity.

Moreover, many features of the federal justice system might promote coercive, mimetic, or normative isomorphism (DiMaggio & Powell, 1983) and militate against local court variation. Perhaps the Guidelines themselves, and the national-level training, guidance, and discourse surrounding them, serve a norm setting function, even post-*Booker*. After close to 20 years of sentencing under the Guidelines, perhaps district courts have come to rely on them to codify and formalize definitions of focal concerns which are incorporated into court communities' culture and sentencing practices. Guidelines would thus be a force encouraging normative isomorphism between courts (DiMaggio & Powell, 1983; for similar arguments regarding state sentencing guidelines see Kramer & Ulmer, 2009; Ulmer, 1997). Furthermore, Main Justice continues to be a potent force for coercive isomorphism and thus Guideline adherence in the practices of local US Attorney's offices, including sentencing recommendations in plea agreements.

In addition, the expectation that local variation and disparity will increase post-Booker assumes that discretion and substantive rationality inevitably carry the cost of unwarranted disparity and that social status characteristics inevitably will color the assessment of focal concerns of sentencing. Recall, however, that later extensions of the focal concerns model do not predict that disparity based on defendant social statuses is inevitable, but rather is context-dependent

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

(Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). If court decision-makers do not link status stereotypes to particular focal concerns, this model would not predict disparity linked to those statuses.

Extending this reasoning, judges inhabit a normative environment that communicates concern about the possibility that social status characteristics might unfairly bias sentencing decisions. Eighteen years of sentencing under the presumptive Guidelines might have established a normative environment of heightened sensitivity to unwarranted disparity. Such a *norm setting* function by the Guidelines might put a brake on any tendency for unwarranted disparity to increase, even if the Guidelines' formal power to constrain sentencing is relaxed. This might be especially true of racial/ethnic disparity, which has been the subject of much legal and policy discourse (especially as related to drug offense sentencing), and around which awareness might be heightened.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Data

The data come from the US Sentencing Commission's Standardized Research Files. Consistent with the USSC's report on the effects of Booker, we utilize four time periods: (1) cases sentenced in the pre-PROTECT Act period which includes fiscal year 2002 (1 October 2001–20 September 2002) and fiscal year 2003 through April of 2003;[3] (2) cases sentenced in the post-PROTECT Act period which includes the second part of fiscal year 2003 (see footnote 3) and fiscal year 2004 through June 2004, which corresponds with the Supreme Court decision in *Blakely v. Washington*[4] (decided 24 June 2004); (3) cases sentenced in the post-Booker period (January 2005 through November 2008);[5] and (4) cases sentenced in the post-*Gall* period (see footnote 5). Consistent with previous research, we exclude all immigration offenses because they are handled differently than other federal crimes (Steffensmeier & Demuth, 2000; Ulmer et al., 2010). Districts with comparatively large numbers of immigration cases commonly employ "fast track" programs designed to expedite such cases (Bowman, 2003). Also, over half of all immigration offenses are for "unlawfully entering or remaining in the U.S." (see §2L1.2 in U.S. Guidelines Manual), with which the US citizens cannot be charged. Therefore, in order to properly compare citizen and non-citizen sentences, these cases should be removed from the analysis. The unit of analysis is each sentenced case, and we limit the universe to cases sentenced in 89 of the nation's 94 federal districts (excluding the US Virgin Islands, Guam, Puerto Rico, the Mariannas, and the District of Columbia).

3. Seven months of fiscal year 2003 were prior to the effective date of the PROTECT Act (1 October 2002–30 April 2003), and five months were after (1 May 2003–30 September 2003).
4. In *Blakely v. Washington*, the Supreme Court invalidated a sentence imposed under Washington State's guidelines, ruling it violated the defendant's rights under the sixth amendment. We remove the period between the *Blakely* and *Booker* decisions to remove any potential *Blakely* effects.
5. The *Booker* decision was decided in January 2005. Therefore, our post-*Booker* period consists of parts of fiscal year 2005, 2006, 2007, and 2008. Only October and November of FY 2008 are included because *Gall* was decided in December, which marks the beginning of our post-*Gall* period.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Dependent Variables

Our analysis examines two dependent variables for each of the four time periods: (1) the likelihood of an imprisonment sentence and (2) the length of sentence. The first dependent variable is a measure of whether an offender was sentenced to a term of incarceration (1 = yes and 0 = no). The second dependent variable is the sentence length ordered for each offender (capped at 470 months).[6] Our dependent variables differ from that used in the USSC 2010 report in that we only use terms of imprisonment in our analysis, whereas their analysis includes months of alternative confinement including home detention, community confinement, and intermittent confinement.[7]

Since the sentence length variable is very positively skewed and regression diagnostics indicated problematic standard errors, we follow previous research and use the natural log of sentence length as our dependent variable. Logging the sentence length, and then re-transforming the regression coefficient by taking its antilog (Hannon, Knapp, & DeFina, 2005)[8] allows us to more substantively examine the proportional differences in sentence lengths associated with our variables of interest. In all tables the untransformed coefficients are presented; however, throughout the text of the paper we discuss the effects from our independent measures using the antilog transformation, which reports the proportional difference in a one-unit increase in our independent measure at the mean of the dependent variable (Agresti & Finlay, 2009; also see footnote 8 for an example).

Independent Variables

Individual level

Consistent with prior research, we control for the *Guideline recommended sentence*. In our sentence length models we include a measure of the presumptive sentence equal to the minimum months of incarceration recommended by

6. We follow the protocol of the USSC and cap the sentence length at 470 months. This is done to "reflect life expectancy of federal criminal defendants more precisely and to provide more accurate length of imprisonment information" (USSC 2009: Appendix A).

7. While it is certainly important to research extralegal disparities in these other forms of confinement, we argue that sentences of home detention (and other forms of confinement) are qualitatively different from time in prison to the point where such sentences should not be analyzed as equivalent forms of incarceration.

8. Researchers in criminology commonly interpret the effects in a semi-logged model directly as a proportional effect. However, this yields results that are somewhat difficult to interpret substantively. For example, a coefficient of 0.05 for black offenders means that blacks receive 5% longer sentences in logged months. Taking the antilog of this coefficient, subtracting 1 and multiplying by 100 $\{([e^{0.05} = 1.0513] - 1) \times 100\}$ yields the proportional difference between whites and blacks in sentence months at the mean of the dependent variable. If the mean sentence length is 60 months, the interpretation is that black offenders receive sentences that are roughly 3.1 months ($60 \times 1.0513 = 63.07$) longer than white offenders.

the sentencing guidelines after adjusting for any mandatory minimum trumps (Albonetti 1998; Engen & Gainey, 2000; Johnson & Betsinger, 2009; USSC, 2004). This measure incorporates the offense severity level, criminal history and accounts for statutory sentencing provisions (i.e., mandatory minimum penalties) that affect the final presumptive sentence. As with sentence length, we cap the presumptive sentence variable at 470 months and take the natural log to reduce positive skewness.[9] In our models of incarceration, we recoded this variable to create a presumptive disposition measure which is coded 1 if the guidelines recommend that the offender be imprisoned and coded 0 if the guidelines recommend probation. All other measures are identical across incarceration and sentence length models.

While criminal history is included in the presumptive sentence measure, we follow previous research (e.g., Albonetti, 1998; Doerner & Demuth, 2010; Johnson & Betsinger, 2009; Johnson et al., 2008; Ulmer, 2005; Ulmer et al., 2010) and include an additional control for the offender's criminal history score.[10] We also control for the type of conviction offense with a set of dummy variables (drug, violent, fraud, firearms, and other offenses, with property offenses as the reference category). We further control for two case characteristics: whether the offender was detained prior to sentencing, coded 1 if the offenders were detained and 0 otherwise and whether the individual was convicted by trial, coded 1 for a trial conviction and 0 otherwise. Our analyses also include dummy variables for whether the defendant received an upward, downward, or substantial assistance (5K1) departure (coded 1 for these departures and 0 otherwise).

Gender is captured with a dummy variable (females = 1), and age with a continuous variable. Race and ethnicity are coded through a series of dummy variables for black non-Hispanic offenders, Hispanic offenders (of any race), and non-Hispanic offenders of "other" races, with white non-Hispanic offenders serving as the reference. We also include a dummy variable for citizenship, with non-citizens coded as 1. Education is captured with four separate dummy variables: less than high school, high school graduates, some college, and college graduates as the reference.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

9. A constant of 0.1 is added to all zero values for the presumptive sentence variable but not the sentence length dependent variable. Taking the log of zero would exclude these values from the analysis. This is appropriate for the dependent variable because we want to analyze only those offenders who actually received a sentence length. The zeros are retained in the presumptive sentence variable (by adding 0.1 to all 0 values) because we want to retain those cases where an offender's minimum sentence was 0 month, but she/he still received a prison sentence.

10. The inclusion of criminal history did not result in problematic collinearity with the presumptive sentence measure (the correlation between logged Guideline minimum and criminal history across the years is .35). Moreover, its inclusion is consistent with previous analyses of federal sentencing decisions (Johnson et al., 2008; Ulmer, 2005; Doerner & Demuth, 2010; Johnson & Betsinger, 2009; Ulmer et al., 2010), and with the USSC's (2006) Booker report. We estimated models where we did not include criminal history with presumptive sentence. The effect for race increased when criminal history was not controlled, but it did so similarly across time periods.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

District level

Several district level (level 2) factors have been found in pre-*Booker* research to affect federal sentencing practices (Anderson & Spohn, 2010; Johnson et al. 2008; Kautt, 2002; Spohn and Fornango 2009; Ulmer et al., 2010). However, the fact that there are only 89 district units of analysis put a premium on parsimony in the inclusion of level-two predictors, so we limited our selection to those factors that would best control for the caseload composition and case processing features of courts. We thus include the following four district-level measures for each time period: (1) district courts' substantial assistance departure rate; (2) district courts' non-substantial assistance downward departure rate; (3) district caseload per judge per year; and (4) district mean final offense levels. The latter variable attempts to control for the average severity of offenses processed by the court, and thus their aggregate caseload's potential to produce severe sentences.

Our research questions necessitate (1) simultaneously examining individual-level and district-level effects, (2) determining whether between-district variation in sentencing outcomes increased post-*Booker*, and (3) whether the inter-district variation in the effects of extralegal variables increased post-Booker. We utilize hierarchical logistic regression models to examine the incarceration decision hierarchical linear models to assess the sentence length decision. The desirability of using hierarchical models for analyses of between-court variation in sentencing and case processing outcomes has by now been established in a number of studies (e.g., Britt, 2000; Fearn, 2005; Johnson, 2006; Kautt, 2002; Kramer & Ulmer, 2009;; Ulmer & Bradley, 2006; Ulmer & Johnson, 2004; Ulmer et al., 2010; Weidner, Frase, & Schultz, 2005; Wooldredge & Thistlethwaite, 2004).

Our hierarchical analysis utilizes random coefficient models, with all individual-level variables centered around their group (district) means, to examine variation both within (i.e., individual level) and between district contexts. This kind of group mean centering in random effects models typically provides comparable results to fixed effects' models (Allison, 2005). In addition, our random coefficient models allow us to examine individual- and district-level variation in sentencing outcomes and between-district variation in individual level predictors of interest (i.e., the effects of extralegal variables).[11]

11. Whether or not analyses of sentencing length outcomes requires controlling for a selection bias for the incarceration decision is a contentious issue in sentencing literature (see Berk, 1983; Bushway, Johnson, & Ann Slocum, 2007). To ensure the robustness of our findings, we ran our sentence length models both with and without a Heckman two-step correction factor, using Stata 10. For the purposes of the Heckman correction, we attempted to find exclusion restrictions and estimate an incarceration (selection) model that was substantively different from the sentence length model (this selection model was also more parsimonious than the in/out model presented). This was difficult, since the majority of the variables that predict imprisonment also predict length. In each way we ran the Heckman correction, the racial and ethnic effects, and those of certain other variables, were greater in size in corrected models in each time period because they affected incarceration decisions more than sentence length decisions. However, the corrected results showed the same patterns in effects across the time periods.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Results

Our analytical strategy proceeds as follows. First, we assess the degree of inter-district variation incarceration and sentence length decisions across our four time periods: pre-PROTECT Act, the PROTECT Act period, post-Booker, and the post-Gall. To do this, we examine descriptive statistics as well as unconditional models that show the amount of variation in incarceration and sentence length decisions that exists between individual cases versus between district courts. Second, we assess the extent to which the effects of predictors of imprisonment and sentence length, especially extralegal variables, have changed across the four time periods. Specifically, we compare coefficients from the four separate time periods using *z*-tests (Clogg, Petkova, & Haritou,. 1995; Paternoster, Brame, Mazzerolle, & Piquero, 1998). Finally, as part of our analysis, we assess the extent to which inter-district variation in the effects of extralegal variables on imprisonment and sentence length has increased by comparing variance components for these variables across time periods.

Table 1 shows the descriptive statistics for the dependent and independent variables for the four time periods. Since the passing of the PROTECT Act the average prison sentence has steadily increased, going from 66 months pre-PROTECT to 72 post-PROTECT, and to about 76 months in the Booker and Gall periods. Moreover, since the Booker and Gall decisions' offenders are slightly more likely to be sentenced to a term of incarceration (87% post-Booker compared to 85% in the periods prior). These trends match the increase in the average presumptive sentence, which went from 64 months to 68, then to 75 and 77 months during this time.

The rate of within-range sentences in the pre-PROTECT era was nearly 66%, which rose to 71% after the passing of the PROTECT Act, and then fell to 60% and 54% in the Booker and Gall periods, respectively. As the rate of within-range sentences dropped, there was a corresponding increase in upward departures and downward departures (non-substantial assistance). Upward departure rates, while still low, increased from 1% pre-PROTECT to 1.8% in the post-Gall era. As might be expected, downward departures decreased after the passing of the PROTECT Act, going from 13% to 9%, only to increase substantially to 20% post-Booker and 25% post-Gall.[12]

Thus, while it appears that judges have taken advantage of their new "freedom" from the guidelines, this does not necessarily mean that uniformity in sentencing decisions has decreased during this time. Therefore, to examine whether this decline in within-range sentences has yielded greater between-district variation, we turn to the results from our unconditional models.

12. This increase is likely overstated because "fast-track" departures and other "government-sponsored" departures from 2003 onward are included within the "other downward departures" category. The USSC did not begin separating downward departures into government sponsored downward departures and other downward departures until 2003.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 1 Descriptive Statistics for all Measures for pre-PROTECT Act, post-*Booker* and post-*Gall* time periods

Dependent Variable	pre-PROTECT	post-PROTECT	post-*Booker*	post-*Gall*
Percent Incarcerated	85%	85%	87%	88%
Mean Sentence length	66.2	71.6	75.9	76.2
Standard deviation	73.3	76.1	77.3	77.4
Independent variables				
Level 1 Predictors				
Offender characteristics				
Mean Age	34.6	34.7	35.4	35.8
White (reference)	38.3%	38.3%	36.7%	36.6%
Black	30.0%	29.9%	31.0%	31.4%
Hispanic	27.8%	27.3%	27.3%	27.0%
Other race	3.9%	4.6%	5.0%	4.9%
Female	15.7%	15.7%	15.0%	14.7%
Less than high school Degree	40.2%	40.6%	39.4%	39.0%
High school graduate	33.9%	34.0%	34.9%	35.0%
Some college	19.2%	19.0%	18.8%	19.2%
College graduate	6.8%	6.4%	6.9%	6.9%
Non-citizen	19.9%	19.2%	19.3%	19.4%
Case characteristics				
Trial	4.1%	5.6%	5.6%	5.0%
Detained pre-sentencing	60.1%	62.7%	66.0%	68.1%
Offense characteristics				
Property (reference)	4.7%	4.2%	3.2%	2.9%
Drug	51.3%	48.7%	48.9%	48.2%
Violent	4.9%	4.6%	4.3%	4.4%
Fraud	22.0%	21.0%	20.3%	19.7%

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 1 (*Continued*)

Dependent Variable	pre-PROTECT	post-PROTECT	post-*Booker*	post-*Gall*
Firearms	10.9%	15.0%	15.8%	15.6%
Other offenses	6.3%	6.6%	7.5%	9.2%
Guideline factors				
Mean criminal history	2.2	2.3	2.4	2.4
Mean presumptive disposition	91%	91%	93%	94%
Mean Presumptive Sentence	63.8	67.5	74.8	76.6
Within range (reference)	65.7%	70.7%	59.9%	54.2%
Upward departure	0.9%	1.0%	1.6%	1.8%
Substantial assistance Dep.	20.7%	19.7%	19.0%	18.8%
Other downward departure	12.8%	8.7%	19.5%	25.2%
Individual *N*	72,035	55,360	143,826	89,440
Level 2 Predictors				
District mean offense severity	19.4	19.6	20.5	20.6
District downward dep. Rate	12%	8%	18%	24%
District Sub. Asst. Dep. Rate	19%	18%	18%	18%
Caseload per Judge (per year)	106.2	104.7	104.1	107.3
District *N*	89	89	89	89

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 2 Unconditional Models of Incarceration (In/Out) and Sentence Length: pre-PROTECT, post-PROTECT, ____ post-*Booker* and post-*Gall*

Incarceration				Sentence length			
pre-PROTECT	Coefficient	Std error	*T*-Ratio	pre-PROTECT	Coefficient	Standard error	*T*-Ratio
Fixed effects				Fixed effects			
Intercept	1.572***	0.054	29.03	Intercept	3.712*	0.031	120.48
Random effects	Variance	Standard deviation	Chi-square	Random effects	Variance	Standard deviation	% Variance
				Level 1 - Individual	1.146	1.070	93.4%
Level 2—district	0.247*	0.497	2764.90	Level 2 - District	0.082*	0.286	6.6%
post-PROTECT	Coefficient	Standard error	*T*-Ratio	post-PROTECT	Coefficient	Standard error	*T*-Ratio
Fixed effects				*Fixed Effects*			
Intercept	1.631*	0.067	24.22	Intercept	3.784*	0.029	131.46
Random effects	Variance	Standard deviation	Chi-square	Random effects	Variance	Standard deviation	% Variance
				Level 1 - Individual	1.138	1.067	94.2%
Level 2 - District	0.384*	0.620	3456.02	Level 2 - District	0.070*	0.265	5.8%
post-*Booker*	Coefficient	Standard error	*T*-Ratio	post-*Booker*	Coefficient	Standard error	*T*-Ratio
Fixed effects				Fixed effects			
Intercept	1.766*	0.060	29.31	Intercept	3.875*	0.026	146.88
Random effects	Variance	Standard deviation	Chi-square	Random effects	Variance	Standard deviation	% Variance
				Level 1 - Individual	1.101	1.049	94.8%
Level 2 - district	0.315*	0.561	6190.86	Level 2 - District	0.061*	0.246	5.2%

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 2 (*Continued*)

Incarceration				Sentence length			
post-*Gall*	Coefficient	Standard error	T-Ratio	post-*Gall*	Coefficient	Std Error	T-Ratio
Fixed effects				Fixed effects			
Intercept	1.794*	0.062	28.81	Intercept	3.884*	0.029	132.69
Random effects	Variance	Standard deviation	Chi-square	Random effects	Variance	Standard deviation	% Variance
				Level 1—individual	1.106	1.051	93.7%
Level 2—district	0.332*	0.576	4956.58	Level 2—district	0.074*	0.272	6.3%

Notes. All models estimated using 89 district courts.
*$p < .001$.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Unconditional Models

Table 2 shows unconditional models (hierarchical models with no predictors) for the four time periods for both the incarceration and the sentence length decisions. In all models there is evidence of significant variation between districts. Important for our purposes, however, is the extent to which Booker and Gall may have changed the degree and nature of this inter-district variation. To examine this question we calculate intraclass correlations for the sentence length models in all time periods which report the amount of between- versus within-district variation.

If courts have become less uniform in their application of the guidelines post-Booker and Gall, we should see an increase in the proportion of the variation in sentence length existing at the district level. However, the results in Table 2 do not show support for those who argue that Booker and Gall will result in decreased uniformity. Indeed, we see that the percentage of the sentence length variance accounted for at the district level is very similar across time periods. While the PROTECT act did appear to reduce the level inter-district variation compared to the pre-PROTECT period (5.8 versus 6.6), do too did the Booker decision (5.2% compared to 5.8% and 6.6%). This is an interesting finding considering that Booker has arguably been the greatest expansion of judicial discretion in the guidelines era (Stith, 2008). This result may suggest that during the uncertain period in the wake of Booker judges were even more reliant on the guidelines. And while Gall appears to have increased the level of inter-district variation slightly (up to 6.3%), it should be noted that this is still less than the level in the pre-PROTECT era. Thus, it appears that inter-district variation in sentencing has changed little after Booker and Gall. If anything, there is slightly less variation between districts in sentencing lengths compared to the pre-PROTECT period. The next sections examine whether extralegal disparities have increased as a result of Booker and Gall and also whether the level of inter-district variation in these disparities has changed over time.

Effects on Incarceration across Time Periods

Table 3A shows hierarchical logistic regression models of whether an offender was incarcerated regressed on offender characteristics, case processing factors, offense categories, Guideline factors, and district-level processes for each of the four time periods. All effects are expressed as odds ratios. The last four columns of the table show results for *z*-tests for the equality of regression coefficients (Paternoster et al., 1998) across time periods.

In nearly every time period there is a consistent pattern of results for extralegal disparities which is consistent with previous federal sentencing research—black and especially Hispanic offenders are more likely to receive incarceration, female offenders are significantly less likely to be imprisoned compared to men, and non-citizens are far more likely to be incarcerated compared to their citizen

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 3A Two-Level HLM models of incarceration in the pre-PROTECT, post-PROTECT, post-*Booker*, and post-*Gall* time periods—individual and district effects

	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
Level 1 predictors	Incarceration		Incarceration		Incarceration		Incarceration						
	OR	SE	*OR*	SE	*OR*	SE	*OR*	SE	z[a]	z	z	z	z
Offender characteristics													
Age	0.991***	0.002	0.992***	0.002	0.991***	0.001	0.991***	0.001	0.176	–0.319	0.105	–0.370	–0.080
Black	1.131*	0.058	1.120	0.065	1.014	0.043	1.137**	0.045	–1.519	–1.284	0.077	0.198	1.858
Hispanic	1.202**	0.061	1.403***	0.066	1.326***	0.061	1.629***	0.074	1.128	–0.624	3.160***	1.502	2.141*
Other race	1.016	0.096	0.948	0.100	1.005	0.071	1.117	0.070	–0.094	0.473	0.795	1.340	1.063
Female	0.611***	0.047	0.624***	0.056	0.570***	0.032	0.551***	0.039	–1.234	–1.414	–1.710	–1.839	–0.681
Less than high school degree	0.987	0.071	0.995	0.072	0.882**	0.047	0.899	0.058	–1.319	–1.409	–1.014	–1.102	0.260
High school graduate	0.875*	0.063	0.843**	0.059	0.881**	0.039	0.816***	0.049	0.083	0.625	–0.891	–0.431	–1.227
Some College	0.915	0.061	0.900	0.061	0.872***	0.035	0.854***	0.046	–0.692	–0.457	–0.900	–0.684	–0.351
Non-Citizen	2.436***	0.102	2.259***	0.107	2.476***	0.090	2.286***	0.091	0.119	0.657	–0.463	0.084	–0.622
Case characteristics													
Trial	4.378***	0.133	3.770***	0.139	4.165***	0.110	4.054***	0.121	–0.289	0.563	–0.428	0.394	–0.166
Detained pre-sentencing	5.743***	0.066	6.588***	0.070	5.853***	0.057	4.898***	0.049	0.217	–1.312	–1.932	–3.484***	–2.377*

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 3A (*Continued*)

	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
Level 1 predictors	Incarceration		Incarceration		Incarceration		Incarceration						
	OR	SE	*OR*	SE	*OR*	SE	*OR*	SE	z[a]	z	z	z	z
Offense characteristics													
Drug	9.097***	0.083	8.318***	0.104	9.910***	0.080	9.189***	0.118	0.740	1.332	0.070	0.634	–0.530
Violent	7.347***	0.104	6.716***	0.171	5.671***	0.146	5.998***	0.136	–1.445	–0.753	–1.183	–0.517	0.281
Fraud	2.346***	0.076	1.933***	0.105	2.363***	0.080	2.710***	0.099	0.065	1.520	1.152	2.338*	1.075
Firearms	3.556***	0.090	2.969***	0.109	3.205***	0.100	3.228***	0.115	–0.773	0.518	–0.664	0.527	0.046
Other offense	1.599***	0.092	1.599***	0.110	1.962***	0.087	2.702***	0.104	1.614	1.458	3.780***	3.468***	2.372*
Guideline factors													
Criminal history	1.482***	0.020	1.431***	0.023	1.464***	0.019	1.441***	0.022	–0.447	0.768	–0.947	0.214	–0.543
Presumptive disposition	29.237***	0.102	34.580***	0.100	42.304***	0.070	46.238***	0.077	2.980**	1.653	3.574***	2.299*	0.855
Upward departure	10.542***	0.324	7.896***	0.305	17.544***	0.393	25.281***	0.379	1.000	1.607	1.754	2.393*	0.670
Substantial assistance departure	0.224***	0.084	0.223***	0.076	0.231***	0.080	0.238***	0.074	0.276	0.322	0.537	0.599	0.258
Other downward departure	0.220***	0.077	0.232***	0.081	0.244***	0.073	0.235***	0.095	0.970	0.459	0.524	0.092	–0.321
Level 2 predictors													
District Mean offense severity	1.263***	0.022	1.270***	0.031	1.262***	0.027	1.251***	0.028	–0.026	–0.146	–0.273	–0.354	–0.224
District downward departure rate	0.983	0.010	0.998	0.017	0.995	0.007	0.987*	0.005	0.959	–0.182	0.374	–0.610	–0.844

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 3A (*Continued*)

	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
Level 2 predictors	Incarceration		Incarceration		Incarceration		Incarceration						
	OR	SE	*OR*	SE	*OR*	SE	*OR*	SE	z[a]	z	z	z	z
District Substantial assistance departure rate	0.982***	0.005	0.974***	0.007	0.976***	0.006	0.972***	0.007	–0.772	0.221	–1.112	–0.188	–0.418
Caseload per Judge (per year)	1.007***	0.001	1.006***	0.002	1.006***	0.001	1.005***	0.001	–0.160	0.383	–0.803	–0.061	–0.591
Intercept	0.222***	0.404	0.253*	0.592	0.267*	0.551	0.406	0.542	0.267	0.066	0.893	0.591	0.544
N	72,035		55,360		143,826		89,440						

*$p < .05$; **$p < .01$; ***$p < .001$

[a]Based on two-tailed test of beta coefficients (not odds ratios).

counterparts. However, the results in last four columns suggest that, with the exception of the findings for Hispanics, which show significantly stronger effects in the post-Gall time period, none of the effects for extralegal variables changed as a result of Booker or Gall. Thus, there is no evidence that overall age, racial, gender, education, or citizenship disparity has increased in incarceration outcomes since the Booker and Gall decisions. There is, however, some evidence that Hispanics are more likely to be incarcerated post-Gall compared with the previous time periods.

Overall, the results from the incarceration analysis are much more notable for their similarities than their differences. In every time period the guideline factors had the greatest effects. Interestingly, one of the only effects that did change in the later time periods was the effect of presumptive disposition, which actually shows a stronger effect in the wake of Booker and Gall. In other words, the recommendation by the guidelines matters most in the post-Gall period at a time when judges were most "liberated" by the constraints of the recommendation.

Finally, the results of our level 2 predictors are consistent across our models. In each time period, districts which had higher mean offense levels and higher caseloads per judge were more likely to imprison offenders. Perhaps unsurprisingly, districts which were most likely to depart downward from the guidelines showed lesser odds of incarceration. However, only the district-level rate of substantial assistance departures showed significant effects in all periods. There were no significant differences in any of the district-level effects, though, across the time periods.

Inter-district Variation in Incarceration Effects

The random effects in Table 3B provide information about the degree to which the effects of select predictors vary across districts. That is, in the analyses in Table 3A, these coefficients were allowed to vary randomly between districts, and the variance components (and their standard deviations) capture the between-district variation in their effects. First, there is evidence of significant variation in effects across district courts for the measures of age, race, ethnicity, gender, and citizenship across at least two of the time periods. However, in only one case the degree of inter-district variation has appeared to have increased substantially—the Hispanic effect. Whereas this effect shows no significant inter-district variation in the pre-PROTECT and post-PROTECT eras, there appears to be a higher degree of variation between districts in the effects of Hispanic ethnicity on imprisonment following the Booker and Gall decisions. For most other effects, however, the results in Table 3B show that there are either similar levels of inter-district variation (age and citizenship follow this pattern) or slightly less variation (race and gender follow this pattern). For example, whereas the race effect shows significant inter-district variation in the first two time periods, the trend has been toward decreasing levels of inter-district variation in the latter two periods. By the post-Gall period, the race

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 3B Two-Level HLM Models of Incarceration in pre-PROTECT, post-PROTECT, post-*Booker*, and post-*Gall* time periods – district level random effects

	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*	
Select measures	Variance	SD	Variance	SD	Variance	SD	Variance	SD
Age	0.000	0.012***	0.000	0.012***	0.000	0.008***	0.000	0.008***
Black	0.104	0.323***	0.133	0.365***	0.059	0.242***	0.036	0.190
Hispanic	0.066	0.256	0.077	0.278	0.135	0.367***	0.160	0.400***
Female	0.087	0.296***	0.107	0.327***	0.027	0.165**	0.045	0.211***
Non-Citizen	0.294	0.542***	0.315	0.561***	0.329	0.573***	0.276	0.525***
Less than high school	0.166	0.407**	0.089	0.298	0.045	0.211	0.095	0.308**
High school graduate	0.113	0.336*	0.046	0.215	0.026	0.162	0.068	0.261
Some college	0.087	0.295	0.058	0.241	0.008	0.089	0.037	0.192
Trial	0.386	0.622	0.338	0.581	0.362	0.602***	0.303	0.551

*$p < .05$; **$p < .01$; ***$p < .001$

effect is relatively stable across district courts. In other words, racial disparity is actually less context-specific following the liberalization of judicial discretion by the Booker and Gall decisions. The weight of the evidence suggests that the differences in sentencing practices across districts that are not attributable to individual-level case factors have either been stable or mattered less in the Booker and Gall eras. Whether similar trends are found for our sentence length analysis is examined in the next section.

Effects on Sentence Length across Time Periods

Table 4A allows us to assess whether extralegal social status-based disparity increased post-Booker and Gall in sentence lengths and whether inter-district variation in such disparity also increased. All independent measures are the same as those in Table 3A with the exception of the presumptive sentence variable, which is now a continuous measure equal to the minimum number of months recommended by the sentencing guidelines (logged).

Similar to the results from our incarceration models there is a consistent pattern of results regarding extralegal disparities—black and Hispanic offenders receive slightly longer prison sentences (roughly 2-4 percent longer sentences), female offenders receive significantly shorter sentences compared to men, education matters little, and slightly longer terms of incarceration are given to non-citizens net of legally relevant controls. Regarding racial disparity, taking the antilog of the coefficients in the models shows that African Americans receive sentences that range from 2% to 4% longer across time periods. However, the length effect for blacks is significantly *smaller* in the post-Booker and Gall periods era compared to the pre-PROTECT period. In other words, our models show that the liberalization of judicial discretion has resulted in significantly less racial disparity compared to the pre-PROTECT era.

Hispanics receive about 3% longer sentences in the pre-PROTECT period, receive sentences that are not significantly different than whites' in the PROTECT era, and receive sentences that are about 3% longer than whites' post-Booker and Gall. However, none of the Hispanic effects are significantly different from each other across time periods. In contrast to the incarceration models, there is no evidence that Hispanic disparity has increased following Booker or Gall.

Consistent with the majority of prior research in state and federal sentencing, women receive significantly shorter sentences than men across time periods. However, the gender effect is actually significantly less in the post-Gall period than in either the pre-PROTECT or PROTECT period. The effects of education on sentence length are negligible in nearly all time periods, and none of these effects appear to have changed over time. Also, there is a slight tendency for non-citizens to receive about 2% longer sentences in the post-Booker and Gall time periods, but this effect is not significantly different across any of the models.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 4A Two-Level HLM Models of Sentence Length in the pre-PROTECT, post-PROTECT, post-*Booker*, and post-*Gall* time periods – individual and district effects

Level 1 Predictors	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
	(Ln) Length		*(Ln) Length*		*(Ln) Length*		*(Ln) Length*						
	b	SE	*b*	SE	*b*	SE	*b*	SE	z^a	*z*	*z*	*z*	*z*
Offender characteristics													
Age	0.000	0.000	0.000	0.000	0.000*	0.000	0.000	0.000	0.887	1.086	–0.127	–0.041	–1.186
Black	0.041***	0.008	0.017*	0.008	0.022***	0.005	0.019**	0.006	–1.976*	0.510	–2.128*	0.207	–0.374
Hispanic	0.030**	0.009	0.006	0.008	0.021***	0.006	0.025**	0.008	–0.809	1.548	–0.374	1.762	0.454
Other Race	0.006	0.020	–0.020	0.018	0.017	0.012	–0.015	0.010	0.487	1.687	–0.973	0.215	–2.143*
Female	–0.125***	0.011	–0.125***	0.013	–0.098***	0.009	–0.077***	0.009	1.866	1.761	3.326***	3.113***	1.690
Less than high school degree	0.014	0.015	0.001	0.015	0.016	0.010	0.001	0.012	0.129	0.840	–0.681	0.001	–0.999
High school graduate	0.015	0.015	–0.002	0.016	0.015	0.009	0.005	0.011	0.029	0.940	–0.511	0.371	–0.690
Some college	–0.003	0.014	–0.007	0.014	0.007	0.009	–0.005	0.010	0.620	0.799	–0.105	0.080	–0.889
Non-Citizen	0.022	0.013	0.015	0.011	0.016*	0.007	0.022*	0.009	–0.418	0.102	–0.004	0.523	0.563
Case Characteristics													
Trial	0.233***	0.025	0.221***	0.022	0.170***	0.009	0.144***	0.011	–2.373*	–2.150*	–3.230**	–3.108***	–1.792
Detained pre-sentencing	0.176***	0.012	0.171***	0.012	0.157***	0.007	0.158***	0.008	–1.363	–0.973	–1.208	–0.834	0.136

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 4A (*Continued*)

Level 1 Predictors	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
	(Ln) Length		*(Ln) Length*		*(Ln) Length*		*(Ln) Length*						
	b	SE	*b*	SE	*b*	SE	*b*	SE	z^a	*z*	*z*	*z*	*z*
Offense characteristics													
Drug	0.251***	0.055	0.274***	0.054	0.143***	0.023	0.052	0.032	−1.826	−2.247*	−3.149***	−3.561***	−2.330*
Violent	0.251***	0.045	0.257***	0.038	0.110***	0.020	0.027	0.028	−2.827**	−3.388***	−4.191***	−4.845***	−2.405*
Fraud	0.000	0.019	0.002	0.025	−0.003	0.020	−0.042	0.027	−0.128	−0.147	−1.276	−1.181	−1.163
Firearms	0.100**	0.036	0.103**	0.034	0.033	0.020	−0.015	0.029	−1.603	−1.784	−2.449*	−2.638**	−1.355
Other offense	0.056*	0.026	0.088**	0.031	0.068***	0.021	0.026	0.032	0.362	−0.534	−0.730	−1.400	−1.110
Guideline Factors													
Criminal history	0.051***	0.005	0.049***	0.004	0.033***	0.002	0.028***	0.002	−3.538***	−3.285***	−4.504***	−4.314***	−1.778
Ln presumptive sentence	0.699***	0.031	0.698***	0.030	0.783***	0.012	0.815***	0.011	2.553*	2.634**	3.562***	3.666***	2.023*
Upward departure	0.761***	0.080	0.829***	0.122	0.711***	0.030	0.648***	0.029	−0.583	−0.936	−1.329	−1.441	−1.513
Substantial assistance departure	−0.534***	0.022	−0.524***	0.025	−0.575***	0.024	−0.588***	0.023	−1.263	−1.474	−1.720	−1.902	−0.407
Other downward departure	−0.490***	0.033	−0.495***	0.032	−0.422***	0.014	−0.420***	0.013	1.871	2.075*	1.970*	2.181*	0.139

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 4A (*Continued*)

Level 2 Predictors	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*		(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
	(Ln) Length		*(Ln) Length*		*(Ln) Length*		*(Ln) Length*						
	b	SE	*b*	SE	*b*	SE	*b*	SE	z^a	*z*	*z*	*z*	*z*
District mean offense severity	0.095***	0.008	0.091***	0.007	0.084***	0.008	0.080***	0.007	–1.032	–0.632	–1.441	–1.027	–0.327
District downward departure rate	0.009***	0.002	–0.012***	0.002	–0.010***	0.001	–0.009***	0.001	–0.109	0.966	0.021	1.100	0.214
District substantial assistance. Departure rate	–0.004*	0.001	–0.004**	0.001	–0.004**	0.001	–0.004**	0.001	–0.364	–0.334	–0.394	–0.367	–0.049
Caseload per judge (per year)	0.000	0.000	0.000	0.000	0.000	0.000	–0.000	0.000	0.984	1.170	–0.328	–0.267	–1.466
Intercept	2.067***	0.164	2.212***	0.140	2.420***	0.189	2.602***	0.168	1.409	0.883	2.277*	1.780	0.717
N	58,708		45,455		119,494		73,986						

*p <.05 **p < .01 ***p < .001

[a]Based on two-tailed test.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

There are interesting post-Booker and Gall differences in the sentencing of particular offenses. While offenders sentenced for drug, violent, firearms, and other offenses all received longer prison terms compared to property offenders both before and after the passing of the PROTECT act, these offense-type differences waned considerably following the Booker decision and have disappeared completely in the post-Gall period. This could arguably be considered evidence of the kind of increased uniformity that the guidelines initially sought to foster because there are no significant sentencing differences among different types of offenders *once relevant legal controls are introduced*. In other words, the severity of the crime and its circumstances as indicated by the Guideline minimum is what appears to matter most, not the type of crime.

In addition, defendants do not appear to be penalized as much for more extensive criminal histories, at least net of the presumptive sentence, in the wake of Booker and Gall. Each increase in the criminal history score was associated with about a 5% increase in sentence length in both pre-Booker periods, but only a 3% increase in the post-Booker and Gall periods.

Some key case processing factors also exhibited interesting differences across time, and these differences do not support the notion that Booker has brought an increased tendency for federal courts to weigh extra-Guideline factors more heavily in sentencing. First, notable "trial penalties" are evident in every time period, which supports prior research on the effects of trial conviction on federal sentencing (Ulmer et al., 2010). However, these trial penalties are notably and significantly smaller in the post-Booker and Gall periods than in either of the two earlier periods (note also that the trial effect is net of the effect of substantial assistance departures or acceptance of responsibility—the latter is captured by the Guideline minimum). Thus, net of Guideline-relevant factors, defendants who go to trial and lose are apparently penalized notably less in the post-Booker and Gall periods than in earlier periods. Furthermore, those detained before sentencing receive significantly longer sentences across time periods, but this tendency has not increased post-Booker.

One final individual-level factor that is noteworthy is the pattern found for the presumptive sentence measure. Similar to the results for presumptive disposition in the incarceration models, the effect of presumptive sentence is considerably and significantly stronger following the Booker and Gall decisions compared to the earlier two time periods. Thus, rather than disregarding the guidelines after Booker and Gall, judges appear to actually be more reliant on their recommendations.

Finally, similar to the incarceration results we find that some of our district-level measures significantly predict sentence length outcomes. Specifically, offenders sentenced in districts with higher mean offense levels receive longer prison sentences net of their own criminal conduct. Offenders sentenced in districts with higher rates of substantial assistance and other downward departures, on the other hand, receive slightly shorter prison sentences. Our caseload measure does not appear to affect offenders' sentence length outcomes.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Table 4B Two-Level HLM models of sentence length in pre-PROTECT, post-PROTECT, post- *Booker*, and post-Gall time periods—district level random effects

	(1) pre-PROTECT		(2) post-PROTECT		(3) post-*Booker*		(4) post-*Gall*	
Select Measures	Variance	SD	Variance	SD	Variance	SD	Variance	SD
Age	0.0000	0.002***	0.0000	0.002***	0.0000	0.0014***	0.0000	0.001***
Black	0.0028	0.053***	0.0021	0.046***	0.0008	0.0281***	0.0017	0.041***
Hispanic	0.0028	0.053***	0.0025	0.049***	0.0009	0.0295***	0.0028	0.053***
Female	0.0052	0.072***	0.0067	0.082***	0.0054	0.0733***	0.0065	0.080***
Non-Citizen	0.0061	0.078***	0.0056	0.075***	0.0022	0.0466***	0.0046	0.068***
Less than high school	0.0078	0.088***	0.0053	0.073***	0.0023	0.0482***	0.0053	0.072***
High school graduate	0.0082	0.091***	0.0064	0.080***	0.0023	0.0482***	0.0053	0.073***
Some college	0.0068	0.082***	0.0041	0.064**	0.0020	0.0449***	0.0038	0.062***
Trial	0.0018	0.043	0.0030	0.055	0.0027	0.0514***	0.0033	0.057***

*$p < .05$ **$p < .01$ ***$p < .001$

Furthermore, none of our district-level effects appear to have changed as a result of Booker or Gall.

Inter-district Variation in Sentence Length Effects

Finally, Table 4B shows the variance components for the random effects of our predictors of interest. Almost all of these effects display significant between-district variation in each time period. That is, the effects of race, ethnicity, gender, etc. vary significantly between districts in each time period. The square root of the variance component provides a between-district standard deviation of the slopes for these predictors of interest. For example, in the pre-PROTECT period, the black–white sentencing differences vary by .053, such that blacks in districts whose race effects are one standard deviation above the mean effect face sentence lengths that are about 10% greater than whites, while in those districts whose race effects are one standard deviation below the mean effect, blacks' sentences are about 1% shorter than whites. In the post-Booker and Gall periods, the standard deviation of the race effect is .028 and .04 respectively, less (but not significantly so) than the inter-district variation in the race effect pre-PROTECT Act. In other words, there is less, not more, variation between districts in the extent to which race influences sentence length. That we find significant variation in the effects of extralegal predictors coincides with several previous multilevel studies of federal sentencing focusing on the pre-PROTECT period (Johnson et al., 2008; Kautt, 2002). However, in no instance do these extralegal effects display significantly more between-district variation in the post-Booker and Gall periods than they do in prior periods. In fact, in separate analyses (not shown) we ran *z*-tests to compare whether the variance components were significantly different across time periods.[13] These results showed that the level of inter-district variation in the effects of extralegal factors had either not changed as a result of Booker and Gall (the most common finding), or had decreased substantially in the post-Booker era compared to the earlier two time periods. We found no evidence that inter-district variation in effects were significantly greater in the Booker and Gall periods. Thus, the findings from Tables 4A and 4B suggest that extralegal effects have not increased in the wake of Booker and Gall, and there does not appear to be increased inter-district variation in the effects of extralegal factors on sentence length.

13. The estimated variance components are assumed to have an asymptotic normal distribution with a standard error in line with the general principles of maximum likelihood. Because the *z*-test is designed to compare two estimates with normal distributions from independent samples (Paternoster et al., 1998; see also Clogg et al., 1995); it is appropriate to use this method to investigate the extent to which the level 2 variance is significantly different across time periods. We calculated the standard errors of the variance components using the full maximum likelihood option in HLM. We thank D. Wayne Osgood for his suggestion that the *z*-test could be used in this way.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Discrepancies with the Findings of the USSC 2010 Report

Finally, our findings are congruent with the USSC's 2006 report on sentencing in the aftermath of Booker, but they contrast with their 2010 report, which finds that black–white disparity has increased in the later post-Booker period, particularly for black males, and that non-citizens are sentenced increasingly harshly in the late post-Booker and Gall periods. For the post-Gall period, their "Booker report model" found a 10% black–white sentence length difference, and their "refined model" found a 21% length difference between black and white males.

A full comparison of our results with the USSC 2010 report is beyond the scope of this article. However, in another paper, we present a replication of the USSC's findings and a set of alternative analyses that make different assumptions and modeling strategies (Ulmer, Light, & Kramer, in press). In general, the differences between our findings and the USSC 2010 report largely stem from our two-stage modeling procedure where we assess incarceration decisions separately from sentence length decisions, whereas the USSC 2010 report combined these decisions into one ordinary least squares (OLS) model. This alternative analysis also presents an extensive discussion on the extent to which disparity in substantial assistance and non-substantial assistance departures has increased post-Booker and Gall. In connection with our multilevel analysis here, we also assessed whether disparity in non-substantial assistance departures, including judge-initiated departures, changed following Booker and Gall, and we examined the degree of inter-district variation extralegal effects in these latter periods. Our findings were generally consistent with those reported above. The weight of the evidence suggests that extralegal disparities in such departures have changed little in the wake of Booker and Gall, and the degree of inter-district variation in these effects is also largely unchanged.

Conclusions

If a primary goal of federal sentencing reform was a reduction of unwarranted disparity, the impact of the Booker decision on extralegal disparity is among the most important empirical questions facing sentencing policy-makers (Hofer, 2007). Indeed, US Attorney General Eric Holder (2009) emphasized the need for such research in recent remarks to Congress. This study answers the call for such research. The aftermath of the Booker decision provides an opportunity to examine what happens when legal decision-makers are released from relatively strong formal rational decision-making constraints, and are given more room to base their decisions on substantively rational criteria. Thus, we partially answer Engen's (2009) call for research that investigates the relationship between changes in sentencing policies and changes in the use of discretion. Does increased freedom to individualize and localize sentencing lead to greater extralegal disparity and greater entropy in federal sentencing (see Richman, 2008)?

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

In general, increased discretion and room for substantive rationality have not resulted in greater extralegal disparity, at least up to FY 2009. Extralegal differences have generally not increased in either the post-Booker or the Gall period, especially compared to the pre-PROTECT Act period. In fact, gender and race differences in sentence length are slightly, but significantly, smaller post-Booker compared to the pre-PROTECT period, and most of the effects of other extralegal characteristics have changed little in the incarceration decision. There is evidence, though, that Hispanic incarceration disparity has increased in the post-Gall period. Furthermore, inter-district variation in sentencing has not increased significantly post-Booker or Gall. Federal district courts vary substantially in terms of the effects of certain extralegal variables on sentencing, but there is no greater variation between districts in these extralegal effects post-Booker and Gall than before.

We are not arguing that there is no extralegal disparity, or that federal district courts do not vary widely in their sentencing patterns, their propensity to deviate from the Guidelines, or the way they allow extralegal and/or non-Guideline factors to influence sentencing. These features of federal sentencing have been established by previous, pre-PROTECT Act research (e.g., Albonetti 1997, 1998; Doerner & Demuth, 2010; Johnson et al., 2008; Kautt, 2002; Mustard, 2001; Steffensmeier & Demuth, 2000; Ulmer et al., 2010) and are confirmed here, including for the post-Booker and post-Gall periods. However, disparity and local variation do not appear to have increased long term in the aftermath of Booker, at least up to FY 2009.

Federal courts do seem to be sentencing particular offenses, especially drug and violent crimes, to shorter prison terms than prior to Booker, even when not deviating from the Guidelines. This is not surprising given the very prevalent view among federal judges and other commentators that the drug Guidelines in particular were too severe. Indeed, a 2006 survey of federal court actors found that 77% of federal judges in the sample felt that current drug guidelines were "too harsh" (Ulmer & Light, 2010). Federal judges may be using their new discretion to depart and "correct" what they see as overly harsh Guideline sentences, much like Pennsylvania judges studied by Kramer and Ulmer (2009) who were faced with increased state guideline sentences for violent offenses.

Federal judges may also be using a variety of other substantively rational criteria to inform their assessments of focal concerns in such departure decisions. For example, Hofer (2007) notes that employment, drug and alcohol dependence, age, family/community ties, and mental and emotional conditions are cited in a larger portion of cases after the Booker decision than they were before. Given our findings, it seems that greater freedom for judges to consider such substantive rationality, and to use their own interpretations of focal concerns in sentencing, needs not result in greater extralegal disparity or greater between-district variation, at least within the confines of an overall, now-advisory Guidelines regime.

An implication of the liberation hypothesis is that disparity will be greater in situations which allow greater discretion. By extension, a major policy change

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

that increases the discretion of judges to individualize and localize sentences should be accompanied by increases in extralegal disparity. Disparity would be the price of allowing greater substantive rationality in decision-making. However, in the wake of the liberation of judges' discretion by Booker and Gall, federal judges seem no more likely than before to consider extralegal factors in their sentencing decisions. Thus, the notion that substantive rationality and extralegal disparity are inextricably linked may be simplistic.

Two sets of reasons might explain why levels of disparity and local variation have not increased post-Booker: (1) a norm setting function of the Guidelines, as they have become embedded in the organizational and legal culture of federal court communities and (2) it is unclear why there would be an *increased* linkage of focal concerns of sentencing to defendant social statuses post-Booker and Gall. First, as Reitz (2005) observed, the Guidelines continue to structure federal sentencing in the aftermath of Booker—courts must continue to calculate them, judges must consider them and must provide legally defensible reasons for deviating from them. In this regard, the federal Guidelines now have a legal status similar to many state sentencing guidelines (Kramer & Ulmer, 2009). Thus, the Guidelines are still likely a force of normative isomorphism (Scott, 2008). A survey of federal judges conducted in 2006 provides some suggestive evidence in this regard. Even after the Booker/Fanfan decision, 56% of federal judges surveyed agreed or strongly agreed with the statement "Congress has spoken and I have sworn to uphold what they have done" as a reason for conforming to the Guidelines (Ulmer & Light, 2010).

This raises the questions of how sentencing guidelines constrain discretion, and whether they need to be mandatory to do so. Evidence exists that a major reason Pennsylvania's guidelines, which were never mandatory, changed sentencing practices and seem to have reduced disparity over time is because they became embedded in local court communities as legitimate and taken for granted decision-making tools (Kramer & Ulmer, 2009; Ulmer, 1997; Ulmer & Kramer, 1998). Likewise, it may be that the Guidelines continue to be the dominant consideration in federal sentencing because they continue to be useful as uncertainty reduction tools. What have federal judges done when constraints on their sentencing discretion were loosened? The evidence here shows that they largely still followed the Guidelines and in fact may be even more reliant on the recommendations from the guidelines as evidenced by the fact that the effect of presumptive sentence post-*Gall* is larger than in all other time periods. Also, 62% of judges surveyed agreed with the statement, "I think the Sentencing Commission has done a pretty remarkable job, when you think about all the different crimes and putting sentences to these crimes" (Ulmer & Light, 2010). These pieces of evidence suggest that judges still regard the Guidelines as useful normative and practical tools for reaching sentencing decisions. Most of the time, it is probably easier for them to follow the Guidelines than not to do so.

Third, the focal concerns model of sentencing argues that social statuses affect sentencing only to the extent that they affect the assessment of blameworthiness, community protection, and practical constraints/consequences. If

federal court decision-makers inhabit a normative environment that communicates concern about the possibility that social status characteristics might unfairly bias sentencing decisions; they may be wary of linking such status characteristics to focal concerns, at least in an obvious and visible manner. This might be especially true, for example, of racial/ethnic disparity, which has been the subject of much legal and policy discourse (especially regarding drug sentences and the crack/powder cocaine disparity, culminating in the Supreme Court's decision in *Kimbrough*), and around which awareness might be heightened. The predictions that extralegal disparity would increase post-Booker would seem to imply that many federal judges exhibit a latent propensity to make defendant social status-based attributions regarding focal concerns, and that this propensity was only held in check by mandatory Guidelines. Likewise, the prediction that inter-district variation would increase post-Booker and Gall assumes that local federal courts exhibit a propensity to ignore sentencing standards set in Washington, DC, and only the restrictions of presumptive Guidelines held this propensity in check. Our findings raise questions about both of these assumptions, at least so far.

Our study is certainly not the last word on the impact(s) of Booker and its aftermath on federal sentencing. For example, previous research has shown that racial and ethnic disparities are more pronounced for drug offenses (Steffensmeier & Demuth, 2000; USSC, 2004) and that the joint effects of several social status characteristics (i.e., young-black-male offenders) are larger than the independent effects of any one of these characteristics alone (Doerner & Demuth, 2010). Future research should evaluate sentencing outcomes post-Booker and Gall for specific types of offenders and specific offenses (e.g., young minority males and drug offenses) pre- and post-Booker.

Relatedly, perhaps the chief limitation in our study is our inability to address disparities that might occur in earlier stages of case processing, such as charging and conviction processes. We framed our research question as evaluating how Booker and Gall affected judges, since the relationship between judicial sentencing discretion and the Guidelines was the focus of Booker and follow-up decisions. However, we also recognize that prosecutors have always played a crucial role in federal sentencing, especially under the Guidelines (see also Bushway & Piehl 2007). Offenders' exposure to Guideline punishments is to a great extent (but not totally) a product of prosecutors' charging decisions and the plea agreement process, in which negotiated stipulations about Guideline-relevant conduct and offense-specific behavior (which raise or lower the final offense level) are commonplace (Ulmer, 2005; Ulmer et al., 2010). It has been long recognized that changes in sentencing schemes affect the distribution of discretion among court actors (Engen, 2009; Reitz, 1998). It may be that federal prosecutors' decisions and behavior in the charging and plea agreement process have changed significantly in the wake of Booker and Gall, and changed in non-uniform ways. Our lack of pre-sentence stage data means that we may be understating overall disparity and inter-district variation due to prosecutors' charging decisions and plea agreement behavior (see Klepper, Nagin, & Tierney, 1983).

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Some 2006 survey data from federal judges, Federal Public Defenders, CJA Panel attorneys are suggestive regarding some ways federal prosecutors might be adapting to the post-Booker environment (Ulmer et al. 2008; Ulmer & Light, 2010). Respondents report three notable post-Booker developments in plea agreements: (1) a 10% increase in the practice of the US Attorneys' offices requiring defendants to waive their rights of appeal, (2) a 10% increase in the prevalence of including stipulations about relevant conduct-/offense-specific behavior and a 2% increase in explicitly specifying the final offense level as part of the plea agreement, and (3) a 12% increase in the frequency with which defense attorneys enter "open" guilty pleas (pleading guilty without a formal plea agreement). On the one hand, the first and second developments suggest that US Attorneys' offices may be more frequently trying to more explicitly structure plea agreements to constrain judges' post-Booker sentencing discretion. On the other hand, the increase in open pleas suggests that defense attorneys may be becoming bolder in avoiding negotiations with prosecutors and relying on the sentencing discretion of judges in hopes of obtaining below-Guideline sentences. A similar use of open pleas by defense attorneys to circumvent prosecutors and gamble on judicial leniency was been found in case processing under Pennsylvania's guidelines (Ulmer, 1997). Future research should explore how prosecutorial and defense practices have adapted to the post-Booker regime.

The Booker era has loosened the restrictions on judicial discretion, and has arguably empowered judges in relation to prosecutors (Stith, 2008). Some proposed remedies would rein in judicial discretion and transform the Guidelines into a system of mandatory minimums. These remedies rest on fears of increased extralegal disparity and between-district variation therein if constraints on judges are relaxed. However, we find little evidence of the increased extralegal disparity or district-level entropy away from Guidelines that many commentators have feared.

Acknowledgments

We thank Shawn Bushway and D. Wayne Osgood for their helpful input on earlier drafts of this paper. This research was supported by a grant from the National Science Foundation, Award no. SES-0350542. An earlier version of this paper was presented at the 2009 American Society of Criminology Annual meeting in Philadelphia.

References

Agresti, A., & Finlay, B. (2009). *Statistical models for the social sciences* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Albonetti, C. A. (1997). Sentencing under the federal sentencing Guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991-1992. *Law & Society Review, 31*, 789–822.

Albonetti, C. A. (1998). The role of gender and departures in the sentencing of defendants convicted of a white collar offense under the federal sentencing Guidelines. In J. T. Ulmer (Ed.), *Sociology of crime, law, and deviance* (vol. 1) (pp. 3-48). Greenwich, CT: JAI Press.

Allison, P. (2005). *Fixed effects regression methods for longitudinal data.* Cary, NC: SAS Publishing.

Anderson, A., & Spohn, C. (2010). Lawlessness in the federal sentencing process: A test for uniformity and consistency in sentencing practices. *Justice Quarterly, 27* (3), 362–393.

Berk, R. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48*, 386–398.

Bowman, F. O., III. (2003, September 23). *Only suckers pay the sticker price: The effect of "Fast Track" programs on the future of the sentencing guidelines as a principled sentencing system.* Testimony before the U.S. Sentencing Commission, Washington, DC.

Britt, C. L. (2000). Social context and racial disparities in punishment decisions. *Justice Quarterly, 17*, 707–732.

Bushway, S., Johnson, B., & Ann Slocum, L. (2007). Is the magic still there? The relevance of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology, 23*, 151–178.

Bushway, S. D., & Piehl, A. M. (2007). Social science research and the legal threat to presumptive sentencing guidelines. *Criminology & Public Policy, 6*, 461–482.

Charlton, P. (2003, September 23). *Charging, plea, and early disposition policies.* Testimony before the U.S. Sentencing Commission, Washington, DC.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology, 100*, 1261–1293.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*, 147–160.

Doerner, J. K., & Demuth, S. (2010). The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in U.S. federal courts. *Justice Quarterly, 27*, 1–27.

Engen, R. (2009). Assessing determinate and presumptive sentencing: Making research relevant. *Criminology and Public Policy, 8*, 323–335.

Engen, R., & Gainey, R. R. (2000). Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: The rules have changed. *Criminology, 38*, 1207–1230.

Engen, R. L., Gainey, R. R., Crutchfield, R. D., & Weis, J. G. (2003). Discretion and disparity under Guidelines: The role of departures and structured sentencing alternatives. *Criminology, 38*, 1207–1230.

Fearn, N. E. (2005). A multilevel analysis of community effects on criminal sentencing. *Justice Quarterly, 22*, 452–487.

Feldmeyer, B., & Ulmer, J. T. (in press). Racial/ethnic threat and federal sentencing. *Journal of Research in Crime and Delinquency.*

Frase, R. (2007). The Apprendi-Blakely cases: Sentencing reform counter-revolution? *Criminology & Public Policy, 6*, 403–432.

Hannon, L., Knapp, P., & DeFina, R. (2005). Racial similarity in the relationship between poverty and homicide rates: Comparing retransformed coefficients. *Social Science Research, 34*, 893–914.

Hartley, R., Maddan, S., & Spohn, C. (2007). Prosecutorial discretion: An examination of substantial assistance departures in federal crack-cocaine and powder cocaine cases. *Justice Quarterly, 24*, 382–407.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Hofer, P. J. (2007). *United States v. Booker* as a natural experiment: Using empirical research to inform the federal sentencing policy debate. *Crime & Public Policy, 6*, 433–460.

Holder, E. (2009, June 24). *Remarks for the Charles Hamilton Houston Institute for Race and Justice and Congressional Black Caucus Symposium "Rethinking Federal Sentencing Policy 25th Anniversary of the Sentencing Reform Act,"* Washington DC.

Johnson, B. D. (2005). Contextual disparities in Guidelines departures. *Criminology, 43*, 761–797.

Johnson, B. D. (2006). The multilevel context of criminal sentencing: Integrating judge and county level influences in the study of courtroom decision making. *Criminology, 44*, 259–298.

Johnson, B. D., & Betsinger, S. (2009). Punishing the "model minority:" Asian-American criminal sentencing outcomes in federal district courts. *Criminology, 47*, 1045–1090.

Johnson, B., Ulmer, J., & Kramer, J. (2008). The social context of Guideline circumvention: The case of federal district courts. *Criminology, 46*, 711–783.

Kalven, H., & Zeisel, H. (1966). *The American Jury*. Boston, MA: Little, Brown.

Kautt, P. M. (2002). Location, location, location: Interdistrict and intercircuit variation in sentencing outcomes for federal drug-trafficking offenses. *Justice Quarterly, 19*, 633–671.

Kautt, P. M. (2009). Heuristic influences over offense seriousness calculations: A multilevel investigation of racial disparity under sentencing guidelines. *Punishment & Society, 11*, 191–218.

Klepper, S., Nagin, D., & Tierney, L. -J. (1983). Discrimination in the criminal justice system: A critical appraisal of the literature. In A. Blumstein, J. Cohen, S. E. Martin, & M. H. Tonry (Eds.), *Research on sentencing: The search for reform* (pp. 55-128). Washington, DC: National Academy Press.

Kramer, J. H. (2009). Mandatory sentencing guidelines: The framing of justice. *Criminology and Public Policy, 8*(2), 313–322.

Kramer, J. H., & Ulmer, J. T. (2009). *Sentencing guidelines: Lessons from Pennsylvania*. Boulder: Lynne Rienner.

Mitchell, O., & MacKenzie, D. (2004). *The Relationship between Race, Ethnicity, and Sentencing Outcomes: A Meta-Analysis of Sentencing Research: Final Report*. Submitted to the National Institute of Justice, Washington, DC.

Mustard, D. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *Journal of Law and Economics, 44*, 285–314.

Paternoster, R., Brame, R., Mazzerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology, 36*, 859–866.

Reitz, K. R. (1998). Modeling discretion in American sentencing systems. *Law and Policy, 20*, 389–428.

Reitz, K. R. (2005). The enforceability of sentencing guidelines. *Stanford Law Review, 58*, 155–174.

Richman, D. (2008). Federal Sentencing in 2007: The Supreme Court holds—The center doesn't. *Yale Law Journal, 117*, 1374–1418.

Richter, J. C. (2008). Deja Vu all over again: How post-booker sentencing threatens equal justice under the law. *Federal Sentencing Reporter, 20*, 340–342.

Savelsberg, J. J. (1992). Law that does not fit society: Sentencing Guidelines as a neoclassical reaction to the dilemmas of substantivized Law. *American Journal of Sociology, 97*, 1346–1381.

Scott, W. R. (2008). *Institutions and organizations: Ideas and interests*. Thousand Oaks, CA: Sage Publications.

Shermer, L. O., & Johnson, B. (2010). Criminal prosecutions: Examining prosecutorial discretion and charge reductions in U.S. federal district courts. *Justice Quarterly, 27* (3), 394–430.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Smith, B., & Damphousse, K. (1998). Terrorism, politics, and punishment: A test of structural-contextual theory and the "liberation hypothesis." *Criminology, 36*, 67–92.

Spohn, C. (2000). Thirty years of sentencing reform: The quest for a racially neutral sentencing process. *Policies, processes and decisions of the criminal justice system: Vol. 3. Criminal Justice 2000.* Washington, DC: U.S. Department of Justice.

Spohn, C., & Cederblom, J. (1991). Race and disparities in sentencing: A test of the liberation hypothesis. *Justice Quarterly, 8*, 305–327.

Spohn, C., & Fornango, R. (2009). U.S. attorneys and substantial assistance departures: Testing for interprosecutor disparity. *Criminology, 47*, 813–846.

Steffensmeier, D., & Demuth, S. (2000). Ethnicity and sentencing outcomes in U.S. federal courts: Who is punished more harshly? *American Sociological Review, 65*, 705–729.

Steffensmeier, 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*, 763–798.

Stith, K. (2008). The arc of the pendulum: Judges, prosecutors, and the exercise of discretion. *Yale Law Journal, 117*, 1420–1497.

Stith, K., & Cabranes, J. (1998). *Fear of judging*. Chicago, IL: University of Chicago Press.

Thompson, S. G. (2006). The Booker Project: The future of federal sentencing. *Houston Law Review, 43*, 269–278.

Ulmer, J. T. (1997). *Social worlds of sentencing: court communities under sentencing guidelines*. Albany, NY: State University of New York Press.

Ulmer, J. T. (2005). The localized uses of federal sentencing guidelines in four U.S. district courts: Evidence of processual order. *Symbolic Interaction, 28*, 255–279.

Ulmer, J. T., & Bradley, M. S. (2006). Variation in trial penalties among serious violent offenses. *Criminology, 44*, 631–670.

Ulmer, J. T., Eisenstein, J., & Johnson, B. D. (2010). Trial penalties in federal sentencing: extra-guidelines factors and district variation. *Justice Quarterly, 27*, 560-592.

Ulmer, J. T., & Johnson, B. (2004). Sentencing in context: A multilevel analysis. *Criminology, 42*, 137–177.

Ulmer, J. T., & Kramer, J. H. (1996). Court communities under sentencing Guidelines: Dilemmas of formal rationality and sentencing disparity. *Criminology, 34*, 383–408.

Ulmer, J. T., & Kramer, J. H. (1998). The use and transformation of formal decision making criteria: sentencing Guidelines, organizational contexts, and case processing strategies. *Social Problems, 45*, 248–267.

Ulmer, J. T., & Light, M. (2010). Federal case processing and sentencing before and after the Booker/Fanfan decision: Little has changed. *Journal of Gender, Race, and Justice, 14*, 143-178.

Ulmer, J. T., Light, M. T., & Kramer, J. H. (in press). Racial disparity in the wake of the Booker/Fanfan decision: An alternative analysis to the USSC's 2010 report. *Criminology and Public Policy*.

USSC. (2004). *Fifteen years of guidelines sentencing: An assessment of how well the federal criminal justice system is achieving the goals of sentencing reform*. Washington, DC: Author.

USSC. (2006). *Final report on the impact of United States v. Booker on federal sentencing*. Washington, DC: Author.

USSC. (2009). *2009 Sourcebook of federal sentencing statistics*. Washington, DC: Author.

USSC. (2010). *Demographic Differences in federal sentencing practices: An update of the Booker Report's multivariate regression analysis*. Washington, DC: Author.

Walker, S. (1993). *Taming the system: The control of discretion in criminal justice*. New York: Oxford University Press.

Weidner, R. R., Frase, R. S., & Schultz, J. S. (2005). The impact of contextual factors on the decision to imprison in large urban jurisdictions: A multilevel analysis. *Crime & Delinquency, 51*, 400–424.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011

Wooldredge, J. (2009). Short versus long term effects of Ohio's switch to more structured sentencing on extralegal disparities in prison sentences in an urban court. *Criminology and Public Policy, 8*, 285–312.

Wooldredge, J., & Thistlethwaite, A. (2004). Bi-level disparities in court dispositions for intimate assaults. *Criminology, 42*, 417–456.

Zatz, M. (2000). The convergence of race, ethnicity, gender, and class on court decision making: Looking toward the 21st century. *Policies, processes and decisions of the criminal justice system: Vol. 3. Criminal Justice 2000.* Washington, DC: U.S. Department of Justice.

Downloaded By: [Ulmer, Jeffery T.][Pennsylvania State University] At: 20:29 7 March 2011