Expanding the Empirical Picture of Federal Sentencing: An Invitation

I. An Invitation
I want to start with an invitation to federal court practitioners to consider partnering with social science researchers as we endeavor to examine and assess the sentencing process in the post-Booker era. This collaboration not only would be a gift to researchers like me but also has implications and potential benefits for the administration of justice in practice. In this essay, I will explain why such teamwork is important by describing what is empirically known about federal sentencing, where the gaps exist in that body of research, and why it is crucially important for sociolegal empirical scholarship, federal sentencing policy, and legal practice to close those gaps.

II. Background
In 1984, the United States Sentencing Commission (U.S.S.C.) was established and charged with developing a set of sentencing guidelines that would help decrease disparities in sentence outcomes across federal jurisdictions and ensure consistency and certainty throughout the criminal justice system. The U.S.S.C. was to come up with a scheme that strictly limited the range of possible outcomes for like offenders, that increased certainty that offenders would be punished, and that increased the penalty severity for certain offender categories. Within three years of this mandate, the Commission had drafted a rigid set of guidelines that was put into effect on November 1, 1987, thereby dramatically transforming the nature of federal sentencing overnight. What had been a system that allowed for significant judicial discretion in sentencing convicted individuals became one that required courts to calculate and impose sentences using a detailed formula based primarily on offense characteristics and offenders’ criminal history.

Part of the U.S.S.C.’s mission was to develop a research program on the federal sentencing process and to maintain data documenting outcomes in all federally sentenced cases. This mandate emerged from a stated commitment to an “empirical approach” to sentencing, which would ideally use research to continually improve the sentencing formulas. Such improvements were meant to minimize irrationality and uncontrolled discretion in the system, and presumably would result in consistency and uniformity in sentencing. The Commission has subsequently amassed and maintained what is arguably the most detailed and complete data set on sentencing outcomes that exists in the United States. These data, which are compiled each amendment year from district court submissions and then cleaned, coded, and disseminated, have since been used by a wide range of researchers, both within the Commission and outside of it, to examine a panoply of research questions about sentencing outcomes and their predictors.

Despite the breadth and detail of this data collection endeavor, the Commission’s use of empirical data to initially develop and then subsequently amend the Guidelines has been critiqued on several grounds. For instance, calculations of average sentences imposed in the pre-Guidelines era for various offenses served as the basis for the initial Guideline ranges but, as several scholars have noted, these calculations left out many key sentencing considerations that shape the final sanction. In particular, the Commission ignored the important role that mitigating individualized background factors played in pre-Guidelines sentencing, and it went on to explicitly mandate their exclusion from consideration. Furthermore, many of the changes to the sentencing Guidelines over the years—which more often than not increased the severity of presumptive sentences—had little or no empirical basis, but rather were directed by Congress.

III. The Empirical Case on Federal Sentencing Under the Guidelines
An even more fundamental problem with the body of empirical research on the Guidelines-era federal sentencing process has emerged: Studies have predominantly—almost exclusively—examined sentencing by starting and ending with the formalized rendering of sentence. Given the nature of the U.S.S.C. data set, most methodological designs in some sense conceptually begin with the final outcome—the imposed sentence in a given case—then look to other coded variables (which, again, are gleaned from district court records, including the probation pre-sentence report) to determine what predicts that final outcome. Thus, the predominant analytic approach has been the use of regression techniques to develop models specifying how, and how much, a range of legal

MONA LYNCH
Professor of Criminology, Law & Society, School of Social Ecology, University of California, Irvine
and extralegal variables shape those formal outcomes. A particular concern of this scholarship has been to document and explain unwarranted sentencing disparities for similar offenders as a function of the offender’s race, gender, class, location, and other such factors.\textsuperscript{3}

This approach has resulted in a wealth of knowledge about the effects of the variables known and recorded by the court on actual sentences, but can only very indirectly illuminate the cause and consequence of the sentencing disparities produced by the less formal, more hidden aspects of criminal case processing. Most significantly, at least in the federal context, under the sentencing Guidelines and mandatory minimums, much of the judicial discretion that existed before the Guidelines has shifted to the U.S. Attorney’s office. The U.S. Attorney’s office is exercised behind closed doors, through plea bargains and negotiations. This point is critical because the vast majority of cases—well over 90 percent—are resolved through negotiated pleas rather than through trial. And because prosecutors can control much of the information about the offense, they have some power to circumvent the Guidelines’ “relevant conduct” provision by selectively releasing such information to probation officers, judges, and others when it serves the negotiated deal.\textsuperscript{5} Moreover, this area is precisely where empirical research has been the most limited.\textsuperscript{7}

So, ironically, despite the existence of the uniquely rich, valuable data resource the U.S.S.C. provides and the wealth of important research that this resource has made possible, very little is known (empirically speaking) about the day-to-day processes that precede formal sentencing in the federal criminal justice system. Indeed, notwithstanding the Commission’s commitment to conducting and using social science research to understand and improve sentencing, almost no empirical research directly and systematically examines those predecessor legal practices that have influenced federal sentencing outcomes: charging decision-making practices and their variation by district; modes and styles of plea bargaining; the interrelationship between the various courtroom workgroups in shaping the adjudication process; and how the strength and type of evidence shapes sentence negotiations. These factors, of course, can explain much variation and disparity in sentence outcomes—a primary concern of the Commission—but analyses of outcome data can only offer a very limited window into their effects.

At the core of this imbalance is the problem of access. Outcome data are readily available to researchers, and the U.S.S.C. research staff is very generous in assisting those who want to use it. District courts themselves are also open to the public, so formal proceedings can be systematically observed and, although cumbersome to access, public court records can be obtained. But social science researchers have generally not been able to gain sustained access to the legal organizations and institutions—such as the pretrial divisions of federal probation, local U.S. Attorney’s offices, and branches of the federal defenders—that are responsible for shaping the final, formal sentence.\textsuperscript{5}

And, it appears, the access to these federal criminal justice organizations has become more restricted over the years. Consequently, a valuable body of social science research on the less visible processes relevant to federal sentencing was conducted in the 1970s and 1980s, both before the Guidelines’ implementation and in its early days,\textsuperscript{9} but the volume of such work has diminished since.

\section*{IV. The Significance of This Deficit at This Moment}

Since their implementation in 1987, the Federal Sentencing Guidelines have been the subject of much controversy on a number of fronts: because of the harsh mandated penalties, particularly for drug offenders; in terms of the U.S.S.C.’s emphasis on sentence uniformity above other justice concerns; due to the degree of restriction that the Guidelines have imposed on judges’ sentencing discretion; and because of their complexity.\textsuperscript{10} Yet despite the criticisms by many within and outside of the federal courts, the Guidelines fundamentally transformed federal criminal justice and came to represent a new paradigm of criminal sentencing. Mixed-methods empirical research conducted during the period of transformation, especially work done by Ilene Nagel and Stephen Schulhofer,\textsuperscript{11} was invaluable in providing a contextualized understanding of the limitations, justice impacts, and unintended consequences of the new Guidelines. This work nicely complemented purely quantitative analyses of outcome data to illuminate, for instance, how the Guidelines were circumvented in practice and why sentence disparities (as a function of defendant demographics and locale) persisted despite the Commission’s efforts.

The set of recent Supreme Court cases that have rendered the Guidelines “effectively advisory”\textsuperscript{14} have opened up the possibility that the transformation in federal sentencing brought about by the Guidelines may be substantially dismantled. Beginning with \textit{Gall v. United States},\textsuperscript{15} the United States Supreme Court has returned some sentencing discretion to federal judges, allowing them to impose any sentence as long as it is consistent with the broad purposes of punishment as outlined by Congress. While the U.S.S.C., and many appellate courts, continued to treat the Guidelines as presumptive and even nearly fully mandatory, by 2007 when the Court ruled in both \textit{Gall v. United States}\textsuperscript{14} and \textit{Kimbrough v. United States},\textsuperscript{16} it became clear that the Court meant for the Guidelines to be just that—guidelines. The Court essentially reiterated its position that the Guidelines are merely advisory and that judges are free to sentence outside of the prescribed Guidelines range on the grounds of policy disagreements, at least in the case of crack cocaine cases. Furthermore, in \textit{Gall}, the Court imposed a standard of review that mandates deference to sentencing judges’ decisions and that effectively frees judges to use an
individualized assessment of a given case, and offender, in deciding whether and how to depart from the Guidelines.

These rulings will potentially reshape federal sentencing in dramatic ways, because judges had been relatively hamstrung from exercising individualized sentencing discretion for more than twenty years. Yet, some sociolegal scholarship would suggest that organizational norms are likely to change more slowly and less dramatically than the formal law itself. Indeed, case outcome norms in different districts would be expected to continue to significantly influence sentencing, whereas the language and mechanisms for negotiating to those normative outcomes might change shape in accordance with policy change. This outcome might especially be expected in jurisdictions that handle large numbers of criminal cases, because going rates for prototypical offenses would be relatively institutionalized as starting points for negotiations. Furthermore, as Michael Nachmanoff and Amy Baron-Evans have pointed out, the complicating factor of statutory mandatory minimums in the federal system means that Booker, Gall, and Kimbrough will have differential impact depending on the case type. U.S. attorney charging decisions, and the nature of plea negotiations.

From a policy standpoint, a complete assessment of the effect of Booker on sentencing processes seems critical. The Commission itself has expressed some alarm about the potential consequences of this changing sentencing landscape, suggesting that unwarranted disparities appear to be increasing in the post-Booker era. Although the Commission’s analysis has been subject to some methodological criticism, its interpretation that more recent outcome data reflect a disparity problem has the very real potential to prompt statutory reform proposals before the full empirical picture of the new era of advisory Guidelines sentencing is known. Because the analyses of this legal change in the federal system—by the Commission and by outside academics and researchers—have thus far primarily worked with the U.S.S.C. sentence outcome data as a starting point, no systematic assessment exists as to whether the quality of justice has changed, for better or worse, as judges are able to expand their scope of consideration to include arguably important individualized factors. Outcome data are simply not sufficient for drawing such conclusions.

Furthermore, as Anne Morrison Piehl and Shawn Bushway have empirically demonstrated in work on state courts, outcome data under highly structured presumptive sentencing systems—such as under the pre-Booker Guidelines—reflect less disparity than actually exists because the charge bargaining processes occur fully outside of the formal court. They advise caution in comparing measured disparities between different kinds of sentencing structures because researchers should “expect to find less measured disparity in studies of highly structured systems with conviction data than in more loosely structured systems even if both systems contain similar amounts of total disparity.” In other words, comparisons of pre- and post-Booker outcome data for levels and types of sentencing disparities without being able to control and account for discretionary processes that happen outside of the court may well be misleading, even fatally flawed, empirically speaking. Consequently, basing policy responses solely on such analyses is similarly flawed if the goal is to have an evidence-based, empirically informed system.

The return of discretion to judges and the renewed opportunity for individualized assessment in federal sentencing is important for federal criminal justice policy; also, significant methodological reasons exist for expanding the empirical approach of federal sentencing. The U.S.S.C. data set, as complete as it is, has some weaknesses that may have dramatic consequences for how analyses are interpreted. For example, each case is treated as independent—no coding is done to indicate co-defendant cases—an approach that violates the assumption of independence in many regression models. As a practical matter, this methodological flaw likely accounts for measurable variance in sentence outcomes given that in multiple defendant cases, such as drug conspiracies, one or more defendants may be given added leniency for testifying against others in the conspiracy. Although some of this variance is accounted for with the available substantial assistance and acceptance of responsibility variables, it does not fully capture the nature of the deal-making process and resulting disparity. Furthermore, in terms of post-Booker analyses, Paul Hofer has very thoughtfully laid out the coding problems—and their implications for interpreting sentencing disparities—in a recent Federal Sentencing Reporter piece.

In addition, important theoretical considerations call for expanding the empirical picture of federal sentencing. First, the line of cases leading up to Booker, Gall, and Kimbrough (e.g., Apprendi v. New Jersey, Blakely v. Washington) may indicate a move away from the kinds of long sentences driven by incapacitation and retribution theories, through constraining state power and reinvigorating defendant rights. Taken together with the developments in the federal sentencing cases at issue here, the question of whether the “punitive turn” in U.S. criminal justice is beginning to reverse, as well as whether the individual offender is reascending as a unit of concern within criminal justice, is ripe for exploration.

On a related note, examining legal processes during periods of transformation—as this era clearly is in the federal courts—can potentially shed light on fundamental questions about the quest for justice. Have local district court actors adapted to the loosening of sentencing constraints in ways that fit their own community’s needs? How do changing norms and practices traverse across different districts, and do these changes happen in regionally specific ways? Does increased disparity in formal outcomes across locales or categories of offenders—if they do indeed exist—necessarily mean less justice? Are those deemed to be similar offenders under the Guidelines’ rubric (which primarily considers only offense seriousness and criminal
history as relevant to the sentence decision) truly similar from a more holistic perspective? These kinds of questions are best examined through qualitative field research that allows social scientists to tease out the complexities of the criminal justice process on the ground.  

Indeed, analogous research has looked at how frontline criminal justice workers reshape law and policy in a number of settings. For instance, Candace McCoy’s now-classic study of the reshaping of plea bargaining practices in California after the passing of Proposition 8 in 1982, which banned plea bargaining in felonies, indicated that the law change merely moved the stage at which bargaining took place (prior to the felony indictment or information being filed). The law change also increased prosecutorial power in plea bargaining in that it created pressure to plea bargain earlier and more often because the opportunity disappeared at later stages.  

Research on the gap between crime control policy shifts toward risk management and their implementation by frontline criminal justice workers also reveals how the translation process from formal law and policy to practice often results in very different outcomes than intended. For instance, the collision of risk management ideals with often competing forces has been illuminated in a number of settings. For instance, Candace McCoy’s now-classic study of the reshaping of plea bargaining practices in California after the passing of Proposition 8 in 1982, which banned plea bargaining in felonies, indicated that the law change merely moved the stage at which bargaining took place (prior to the felony indictment or information being filed). The law change also increased prosecutorial power in plea bargaining in that it created pressure to plea bargain earlier and more often because the opportunity disappeared at later stages.  

In light of the Booker, Gall, and Kimbrough cases, now is an opportune time to revisit some of the larger underlying questions about law on the books versus law in action as it relates to federal sentencing and, specifically, about the mechanisms of legal policy change and its implementation. In the year immediately following Booker, data indicated that district courts’ sentencing practices had not changed dramatically from the immediate pre-Booker period.  

Getting a full picture of whether and how Booker is changing sentencing and, if so, whether the outcome data fully reflect the entire justice process that leads to a given sentence, will be crucial for remedying injustices and optimizing fairness in federal court. In other words, looking beyond sentence outcomes to understand how cases get to that final place is necessary for good policy, legal practice, and academic scholarship. And, as federal sentencing gets transformed by a shift of discretionary power back to judges, there is no better time to undertake more contextualized, qualitatively rich examinations.  

So call me. Let’s talk. . . .

Notes


6. On these issues, see Bennett L. Gershman, The New Prosecutors, 53 U. Pitt. L. Rev. 393 (1991); Gerald W. Haeany, The Reality of Guidelines Sentencing, 44 St. Louis U. L.J. 293 (2000); Lauren O’Neill Shermier & Brian D. Johnson, Criminal Prosecutions: Examining Prosecutorial Discretion and Charging Decisions in U.S. Federal District Courts: Who Is Punished More Harshly?, 65 Am Soc. Rev. 705 (2000). In general, the findings indicate that unwarranted disparities have persisted under the Guidelines. Although some forms of disparity have been reduced, the finding that race, ethnicity, and gender continued to have some predictive power in terms of sentence severity has been the biggest puzzle in the Federal Guidelines sentencing scholarship, because it seems to go exactly against the Guidelines’ very engineering.

7. See Anne Morrison Piehl & Shawn Bushway, Measuring and Explaining Charge Bargaining, 23 J. Quantitative Criminology 105 (2007) for an excellent empirical example of how much the role of prosecutorial discretion in bargaining distorts the disparity picture under presumptive guidelines systems.


Ilene Nagel & Stephen Schulhofer, supra note 9.


The analysis amalgamates offense types, which may have led to an imprecise measuring of the level of sentencing disparities and the level of change by offense type, because the larger categories of sentenced offense types diverged in how they affect different demographic groups. Another major question has to do with whether fast-track immigration cases were excluded, thereby leaving in only the more serious or challenging immigration cases.

Pehl & Bushway, supra note 7.

Id. at 122.

See Brian Johnson et al., The Social Context of Guidelines Circumvention: The Case of Federal District Courts, 46 CRIMINOLOGY 737 (2008), for more on this topic.


Simon Hallsworth, Rethinking the Punitive Turn, 2 Punishment & Soc’y 145 (2000).

This work is precisely what has been so scarce over the past twenty years. The exceptions include Jeffery Ulmer’s theoretically rich qualitative work in four federal district courts, before Booker: Jeffery Ulmer, The Localized Uses of Federal Sentencing Guidelines in Four U.S. District Courts: Evidence of Processual Order, 28 SYMBOLIC INTERACTION 255 (2005); Lisa Miller and James Eisenstein’s work on the nexus between federal and state prosecutions in gun and drug cases: Lisa Miller & James Eisenstein, The Federal/State Criminal Prosecution Nexus: A Case Study in Cooperation and Discretion, 30 L. & Soc’y Inquiry 239 (2005); and Nagel & Shulhofer, supra note 9.

