The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Legal change and sentencing norms in federal court: An examination of the impact of the *Booker, Gall,* and *Kimbrough* decisions

Mona Lynch (Principal Investigator)
Department of Criminology, Law and Society
University of California, Irvine
Irvine, CA 92697-7080
(949) 824-0047
lynchm@uci.edu

Marisa Omori (Graduate Research Assistant)
Department of Criminology, Law and Society
University of California, Irvine
Irvine, CA 92697-7080
omorim@uci.edu

ABSTRACT

1. *Research Goals and Objectives:* This funded research project examines how federal drug trafficking cases are sentenced under a rapidly changing federal justice system. The federal sentencing guidelines have lost their authoritative force since the US Supreme Court in *US v. Booker* (2005), *Gall v. US* (2007) and *Kimbrough v. US* (2007) declared that the guidelines are merely advisory in determining criminal sentences in federal court. As a result of these decisions, judges are now free to impose sentences other than those prescribed by the Guidelines, so long as they explicitly justify and explain their reasons for the variance. This research is designed to look specifically at how, and to what extent, federal sentencing has changed over time due to the *Booker, Gall,* and *Kimbrough* decisions, as well as prior policy reforms. It is particularly focused on change at the district level, and examines sentencing practices over time in the use of Guidelines sentencing as well as mandatory minimum sentencing (which was not directly impacted by the *Booker* line of cases).

2. *Research Design and Methodology:* We conduct quantitative analyses of US Sentencing Commission individual sentence outcome data, supplemented with district level and state level variables, to assess sentence outcome variation nationally, by district, for 5 categories of drug trafficking offense types (crack cocaine, powder cocaine, heroin, methamphetamine, and marijuana) subject to section 2D1 of the Guidelines, from 1993 to 2009. Research questions about the quality, variation, and extent of sentencing change over time were tested using hierarchical linear modeling, in the first set of analyses, and a time series cross-sectional (TSCS) design, in the second set of analyses, in order to examine the intersection of time and place in the federal sentencing of drug trafficking.

3. *Data Analysis and Products:* The findings suggest that in any given year, individual case factors explain that bulk of variance in sentence outcome; however, when examined over time, district- (and state-)level factors explain a significant share of sentencing variance. Thus, districts vary considerably from each other over time, but they look like themselves over time, suggesting that local court norms and contexts play an important role in

1

sentencing outcomes. Moreover, districts with high caseload volumes and a high proportion of drug trafficking cases are especially likely to maintain stability in practice over time and account for sentence outcome variance. Findings suggest that the federal system should not be treated as a single, unified system that changes, lockstep, in response to policy mandates. Findings also suggest that in the case of drug trafficking, sentence lengths have steadily decreased over time, as reflected in the difference between the Guideline minimum sentence and actual imposed sentence. This may indicate that, in practice, court actors have corrected for the very long, draconian federal drug sentences that have been passed by Congress and devised by the US Sentencing Commission.

TABLE OF CONTENTS

Executive Summary

Introduction and background to the research problem:

The authoritative force of Federal Sentencing Guidelines has been diminished by a series of legal mandates in recent years. Specifically, the U.S. Supreme Court in *U.S. v. Booker* (2005), *Gall v. U.S.* (2007), and *Kimbrough v. U.S.* (2007) ruled that the Guidelines are merely advisory in determining criminal sentences in federal court. As a result of these decisions, judges are now free to impose sentences other than those prescribed by the guidelines, so long as they explicitly justify and explain their reasons for the variance. These rulings have had the potential to dramatically reshape federal sentencing, as judges had been relatively hamstrung from exercising individualized sentencing discretion since the Guidelines were put into effect in 1987.

Yet, some socio-legal scholarship would suggest that organizational norms are likely to change more slowly and less dramatically than the formal law itself. Indeed, it is expected that case outcome "norms" would be relatively resistant to change, while the language and mechanisms for negotiating to those normative outcomes might change shape. Moreover, the way that federal district courts adapt to this transformed legal landscape is likely to vary considerably by locale, so its impact may diverge in intensity, quality, mechanisms, and outcomes. This NIJ-funded research was designed to examine whether and in what ways these formal legal changes have trickled down into federal sentencing practice.

The project is comprised of two key dimensions—time and place. The analyses track changes in sentencing outcomes (and, less directly, sentencing processes) over time in an effort to measure how these court decisions, along with other sentencing policy changes, affect sentencing behavior on the ground. Integrated into the conceptual model is the notion that place also matters—both in terms of local legal structures and norms, as well as broader socio-political cultural norms. The analyses are limited to U.S. Sentencing Commission (USSC) federal drug trafficking case outcome data from fiscal years 1993-2009.[1] Drug trafficking cases were the focus for several reasons. Drug trafficking was, until 2008, the single largest annual category of federally sentenced offense since the

---

[1] We also obtained data for fiscal year 1992, but we were unable to use it because it did not including coding for mandatory minimum sentences, which is one of our key variables of interest.

inception of the Guidelines, so it comprises a significant share of federal court interest and resources. Drug trafficking sentences also grew dramatically more punitive under the Guidelines, more so than any other category of offense. In addition, drug trafficking offenses are especially likely to be subject to mandatory minimum sentences, so provide a comparative measure of change since 2005, since mandatory minimums were not affected by the *Booker* line of cases.

This research project builds upon the insights of a large body of existing research that examines how the Guidelines have been operationalized since their inception. A particular interest of prior research has been to assess whether and how much the Guidelines have achieved their goal of ensuring uniformity in sentencing across like offenders. Thus, a number of scholars have looked at how much and under what conditions "extra-legal" defendant and contextual factors such as race, ethnicity, class, gender, and geography impact sentence outcomes (e.g., Albonetti, 1997; Everett and Wojtkiewicz, 2002; Kautt, 2002; Mustard, 2001; Steffensmeier & DeMuth, 2000; Wu and Spohn, 2010). Findings generally indicate that although some forms of disparity were reduced with the implementation of the mandatory Guidelines, race, ethnicity, and gender continued to have some predictive power in terms of sentence severity, as have locale and other district-level variables (e.g. Doerner and Demuth, 2010; Feldmeyer and Ulmer, 2011; Ward, Farrell and Rousseau, 2009; Wu and Spohn, 2010).

Research by Johnson, Ulmer, and Kramer (2008), using pre-*Booker* sentencing data, helps explain the persistence of disparities under the mandatory Guidelines despite the Commission's efforts to achieve uniformity. Their analysis suggests that Guidelines have been sidestepped in a manner that likely serves local needs and prerogatives, and that fits with local organizational norms. It also indicates that the structure and process of plea bargain negotiations under the specific constraints of the Guidelines, and the substantial discretion held by the U.S. Attorney's offices, are key factors in sentence outcome variations for similar offenders.

An emerging body of empirical scholarship has also looked at the time dimension, with an eye toward how the *Booker* line of cases has changed sentencing outcomes. Jeffery Ulmer and his colleagues (Ulmer, Light and Kramer, 2011a & b) have conducted the most extensive analyses of sentencing change brought on by *Booker*, et al. Their work raises

questions about the USSC's finding of increased racial, ethnic and gender disparities post-*Booker*, demonstrating that while sentence variations between districts, and disparities between similarly situated offenders at the individual level persist, neither has been exacerbated by *Booker*'s or *Gall*'s mandates. Moreover, they found that sentencing disparities as a function of offender demographics in the post-*Booker* period continue to be, in large part, a product of prosecutorial behavior (Ulmer and Light, 2011). Taken together, prior research on federal sentencing suggests the importance of local level factors and court workgroup norms in explaining sentencing in practice.

Methods:

In recognition of the contextually-embedded nature of individual case sentencing, this project focuses in particular on the role of district-level sentencing practices over time in drug trafficking cases. Multi-level quantitative analyses of U.S. Sentencing Commission individual sentence outcome data, coupled with district level and state level variables, were used to assess sentence outcomes nationally, by district, for 5 categories of drug trafficking offense types (crack cocaine, powder cocaine, heroin, methamphetamine, and marijuana) subject to section 2D1 of the Guidelines, for the years 1993-2009. Research questions about the quality, variation, and extent of sentencing change over time were tested using several different analytic methods in order to examine the intersection of time and place in the federal sentencing of drug trafficking. Specific research questions are as follows:

1. Have federal sentences in drug trafficking cases increasingly varied from the presumptive Guidelines sentence recommendations from the pre-*Booker* period through to the post-*Kimbrough/Gall* period?
2. If so, are the increases significantly greater for cases that are not subject to mandatory minimums than those that are?
3. Do districts that handle large numbers of drug cases demonstrate less variation over time compared to districts that handle relatively fewer drug cases, as might be expected by theories about organizational norms and resistance to change? Similarly, do smaller districts demonstrate more variation across all time periods than larger districts, reflecting both a longer-standing, more individualized approach and a decreased reliance on expectations about "going rates" for prototypical cases?
4. What are the qualities of the variations: are they generally increasing or decreasing sentence outcomes for similar offenders over time?

The US Sentencing Commission's annually-assembled dataset, "Defendants Sentenced Under the Sentencing Reform Act" which contains data for all offenders sentenced in federal court in those fiscal years were obtained. These data include information on individual sentenced cases, such as defendant demographics, offense characteristics, case processing characteristics, and final sentence imposed. Drug trafficking cases were identified by using the sentencing guideline offense, yielding a sample of 376,637 cases. Because variation at the federal district and state levels was expected, these data were merged with federal district-level indicators for the 89 non-territorial federal districts from the Federal Court Management Statistics website, and state level demographic data from the U.S. Census Bureau. Analyses were conducted on both the sample of individual offenders using multilevel regression models, as well as on the 89 districts over time using a time series cross-sectional (TSCS) design.

Because the primary goal of this research is to explore changes in the legal process, three distinct sentencing outcomes were specified as dependent variables:

- final prison *sentence length* in months.
- the *sentence difference,* calculated as the difference between the Guidelines minimum and actual sentence.
- Whether or not a *mandatory minimum sentence* was imposed.

Independent variables of interest are the *Booker* case, decided on January 11, 2005, and *Gall* and *Kimbrough* cases on December 9, 2007. Additional time periods were also specified since each of the specified periods represents distinct policy mandates related to federal sentencing (see Ulmer, et. al, 2011a for more details). The full range of time periods are pre-*Koon* (before June 1996), *Koon* (from June 1996-February 2003), and the PROTECT Act (March 2003-December 2004), *Booker* (January 2005-December 2007), and *Kimbrough* (January, 2008 on). The PROTECT Act period (which serves as our reference category in our models) was significant in particular because it represents a period when judges had the least amount of discretion to vary from the Guidelines, and when district-level prosecutors were discouraged from settling cases through discounted sentencing offers. Approximately 19% of the cases in the sample occurred during the *Booker,* and an additional 12% occurred during the *Gall/Kimbrough* time period.

7

A number of individual case characteristics were also included in the multilevel models. Because sentences vary substantially by drug, the primary drug of the Guideline offense--crack cocaine, powder cocaine, heroin, marijuana, methamphetamine, or other drug--was controlled for in the models. The offender's criminal history category, which ranges from 1-6, sentencing adjustments including a variables for the application of a mandatory minimum drug sentence, safety valve application, which allow low-level, first time drug offenders relief from mandatory minimums, substantial assistance and other types of Guidelines-authorized departures were also included. Additionally, offenders' demographic information, such as gender, age, educational attainment, race/ethnicity, number of children parented by the defendant, and U.S. citizenship status, was included.

Districts were expected to respond differently based on local legal norms and case volume, so a set of independent variables at the district level were included in the models: the per-judge criminal caseload size, measured by the number of criminal case filings per judgeship, was included, and the district's relative efficiency was captured by the median time to disposition in months. The percent of the caseload comprised of drug trafficking cases, the percent that were crack cocaine cases, and the percent of Black defendants in the district were also calculated and included. Also included was the state drug treatment admission rate (derived from the Treatment Episode Data Set) and the violent crime rate (derived from the Uniform Crime Reports).

Finally, demographic contextual factors were expected to influence sentencing practices, so a number of state-level demographic variables from the Census Bureau were calculated and included: percent in poverty, percent high school graduate, density of population (per square mile), and percent Black population. These were included in both the multi-level and TSCS analyses. For the TSCS models, a variable specifying geographic regions in the U.S (Northeast, Midwest, West, and South) was also included.

Main Findings:

The findings of this research suggest that national-level sentencing policy changes, including the reforms mandated by the *Booker, Kimbrough,* and *Gall* cases, neither uniformly nor dramatically transformed sentencing practice. First, individual case factors are the single largest predictor of sentence outcome over all the time periods. Second,

sentencing behavior across districts has changed incrementally over time, and did not dramatically shift at each major policy change (see Figures 3-5). Third, when looking at the interaction of time and locale, districts demonstrate a tendency toward consistency in sentence practices over time, and tend to distinguish themselves from each other over time.

Specifically, in any given year, only 6.6% of the variance in sentence length was explained by district level factors, but when comparing sentence length of districts over the years, 65% of the variance in sentence length is between districts or states. In other words, while there is still more variation in sentence length between individual cases than between districts and states generally, there is far greater variation in mean sentence length between districts (and states) than within districts over time. This general pattern holds for the sentence difference and application of mandatory minimums analyses as well. For individual case outcomes, the variance occurs predominantly at the individual case level compared to the district and state in general. When examining the mean sentence difference and the proportion of mandatory minimums in districts over time, however, the proportion of variance explained at the district and state are larger than within districts over time. The estimated variance at the district and state level for mean sentence difference is about 54%, and proportion of mandatory minimums applied is 56%. This suggests that there is considerable between-district (and state) differences in legal outcomes, so district level factors are important predictors of sentencing practices over time.

One of the main questions of interest was whether the loosening of the mandatory nature of the sentencing guidelines as a result of *Booker* and *Kimbrough/Gall* predicted sentencing changes. In general, it appeared that all periods other than the PROTECT Act period were similar to each other in sentencing outcomes. Overall drug trafficking sentences were shorter both before and after the PROTECT Act period, and mandatory minimums were used in a higher proportion of cases in the 7 year period (*Koon*) before the PROTECT Act, and in the *Booker* and *Kimbrough* periods. Overall, there was a counterintuitive finding that the use of mandatory minimums was positively related to deviations from the Guideline minimums. This suggests that prosecutorial behavior is especially influential in sentence deviations, which was especially the case in the years before the PROTECT Act. Additionally, contrary to what might be expected, sentence deviations in the *Booker* and *Kimbrough* periods are no more likely in non-mandatory

9

minimum Guidelines cases than in mandatory minimum cases. Thus, it does not appear the new-found judicial discretion in non-mandatory minimum cases is directly responsible for increases in sentence deviations post-*Booker*.

Measured differences from the guideline minimum sentence steadily increased over time, although the nature of those increases appear to have changed from pre-PROTECT Act to post-*Booker*. This is reflected in the multi-level models measuring the interactive effect of mandatory minimums x time period on the sentence difference variable. Specifically, in the years before the PROTECT Act was instituted, increased application of mandatory minimums was associated with greater deviations down from the guideline minimum. As noted above, this effect disappeared after *Booker*. This finding suggests that the actual mechanisms for achieving desired sentence outcomes changed as a function of policy. Thus, prior to 2005, US Attorneys held the most discretionary power to seek particular sentence outcomes, so sentence discounts in both mandatory minimum cases and regular Guidelines cases were primarily authorized by prosecutors. Once some discretionary sentencing power in Guidelines cases shifted to judges in the wake of Booker, prosecutors likely use mandatory minimums more strategically to obtain particular outcomes, so are less willing to discount those sentences.

When looking at district level effects, an increased number of filings per judgeship predicted shorter sentences, which makes sense if higher-volume areas feel more taxed for resources and consequently have a lower "going rate" for sentences (Ulmer and Johnson, 2004). Moreover, those districts with the highest volume of cases, and those with the highest percentage of drug trafficking cases were the most stable over time in terms of sentencing practices and outcomes, and least influenced by the national policy changes. Specifically, 34% between-district sentence length variation in high case-volume districts was explained by the district, whereas only 11% was explained in medium case-volume districts, and 4% in low case-volume districts.

As was found in the multi-level analyses, in the district-level time series models, the number of filings per judgeship had a small but significant negative effect on mean sentence length, so in districts where caseload size was higher, mean sentence lengths were lower. There were some regional effects as well. All regions have lower mean sentence lengths for trafficking than the South, although the Midwest is not significantly different

from the South in the final model. Compared to the South, all other regions were also significantly more likely to discount final sentences. Thus, it appears that the districts in the South more closely hew to the Guidelines and mete out more punitive sentences than their counterparts in the rest of the nation.

The distinct set of time series cross-sectional models also examined whether the different periods were associated with variations in the use of mandatory minimum sentences. Unlike the findings from the individual-level case models, there are no significant effects over time, suggesting that while there may have been some variation in the use of mandatory minimums over time, there was not a large enough effect to detect it at just the district level. Interestingly, while the drug type results converge to some degree with the individual case models, districts with higher proportions of methamphetamine cases predict higher proportion of mandatory minimums relative to districts with higher proportions of crack cocaine. In the individual case level analyses, crack cocaine trafficking cases have higher odds of being sentenced under a mandatory minimum compared to all other drug types.

These somewhat contradictory findings at different levels of aggregation could reflect the highly regional nature of drug prosecutions at the district level. This is supported by the finding on the effect of region; judicial districts in Midwestern states had significantly higher proportion of mandatory minimums relative to the South, and the Northeast had significantly lower rates. Thus, those districts that pursue the highest percentage of crack cocaine and methamphetamine cases were more likely to use mandatory minimums. In addition, both higher number of filings per judgeship and longer median time to disposition were related to higher rates of mandatory minimum use in districts. In this analysis, higher percentage Black population in the state is also significantly related to a higher proportion of mandatory minimums.

Overall, the findings suggest that districts vary considerably from each other in sentencing over time, so the federal system should not be treated as a single, unified system that changes, lockstep, in response to policy mandates. As Johnson, Ulmer and Kramer (2008) have suggested, district courts reflect local community concerns, district level prosecutors' priorities, and local courtroom workgroup norms, in both caseload qualities and outcomes. The between-district variability begins with the composition of the drug

11

caseload, where some districts handle relatively few trafficking cases, and others have criminal caseloads in which the majority of cases involve drug trafficking. Within that variability is another layer of variation as to the specific types of illicit substances that are prioritized.

Second, the data reflect relative stability of outcomes within districts over time, rather than radical change, including in response to the Supreme Court's mandates in *Booker* and *Kimbrough/Gall*. Drug trafficking sentencing is complicated by high, and highly variable rates of mandatory minimum usage by districts, which again is partly driven by specific drug caseload compositions at the district level. Powder cocaine, crack cocaine, and methamphetamine all are sentenced under mandatory minimums in about 75-80% of trafficking cases, whereas heroin has slightly lower rates, and marijuana's rate is much lower. Consequently, while across all districts, approximately two-thirds of drug trafficking cases are sentenced under mandatory minimums, this rate ranges from a low of about one-third of the trafficking cases in the Southern West Virginia district to 88.4% in the Southern Iowa district. Crack cocaine in particular is heavily enforced in the South, whereas marijuana prosecutions are concentrated along the southwest border states, and methamphetamine is concentrated in the mountain west and parts of the Midwest.

Third, in concurrence with Johnson, et al. (2008), the findings of this study also highlight the importance of considering how sentencing is shaped by a number of meso- and macro-level factors that mediate how policy is put into practice. Thus, sentencing trends transcend policy change, and it appears that while there is some stasis over time in regard to sentence length (with the possible exception of the PROTECT Act period which had longer sentences than all of the other periods), the mechanisms for getting to those outcomes adapt to policy mandates. Notably, because prosecutors' tools for decreasing final sentence outcome are subsumed within the Guidelines minimums (especially the granting of substantial assistance departures), there is less observed variation in our sentence difference measure pre-*Booker* than post-*Booker,* but overall outcomes remain fairly stable. Moreover, in those districts with high volumes of cases in general, and high proportion of drug trafficking cases, we see a strong tendency toward sentencing norms that persist across time periods, suggesting that the local legal culture is an especially

important element to case outcomes above and beyond national level policy mandates in districts that handle large, and drug trafficking-heavy caseloads.

Our findings also call into question the U.S. Sentencing Commission's interpretation of post-*Booker* sentence outcomes, and suggest the importance of using appropriate variable specification, multi-level predictors and controls, and the fullest range of time periods available. We found little evidence that in drug trafficking cases, judges' recently-restored freedom to sentence outside of the Guidelines accounts for increases in observed variations in sentencing at the individual or district levels. In part, this is confirmed by our main effect findings that a higher proportion of mandatory minimum cases predicts more variance from the guideline minimum, and that Guidelines cases are no more likely to have sentence deviations post-*Booker* than are mandatory minimums. Since judges have, throughout the entire period, been constrained from deviating from mandatory minimums, this also suggests prosecutorial behavior is a very important component to such variances. For policy-makers, this would indicate that reinstating constraints on the judicial sentencing process will not achieve desired uniformity. Thus, enacting additional statutory minimum sentences is not an appropriate policy response in that regard, nor would be other versions that aim to constrain judges. Moreover, the observed regional effects transcend jurisdictional lines, suggesting an important influence of broader cultural norms and historical contingencies in how federal courts operate on the ground, which is not easily constrained by national-level policy mandates.

Finally, from a policy standpoint, the existence of a dual system—a Guidelines system that is designed to take into account a large number of case and offender factors so the final calculation provides for a tailored sentence that still aims for some uniformity, along with a mandatory minimums system that by design ignores many of those key sentencing factors carefully crafted by the U.S. Sentencing Commission and is quite blunt and imprecise---appears to allow for an unintended, but nonetheless exploited, layer of prosecutorial discretion in the federal administration of justice. Thus, our findings, among others, at least indirectly suggest that disparities in the use of mandatory minimums and their corollary mechanisms (such as substantial assistance departures) may be more problematic than changes to the mandatory status of the Guidelines.

Introduction:

A. Statement of the Problem:

The authoritative force of Federal Sentencing Guidelines has been diminished by a series of legal mandates in recent years. Specifically, the U.S. Supreme Court in *U.S. v. Booker* (2005), *Gall v. U.S.* (2007), and *Kimbrough v. U.S.* (2007) ruled that the Guidelines are merely advisory in determining criminal sentences in federal court. As a result of these decisions, judges are now free to impose sentences other than those prescribed by the guidelines, so long as they explicitly justify and explain their reasons for the variance. These rulings have had the potential to significantly reshape federal sentencing, as judges had been relatively hamstrung from exercising individualized sentencing discretion since the Guidelines were put into effect in 1987.

Yet, some socio-legal scholarship would suggest that organizational norms are likely to change more slowly and less dramatically than the formal law itself. Indeed, we would expect that case outcome "norms" would be relatively resistant to change, while the language and mechanisms for negotiating to those normative outcomes might change shape. Moreover, the way that federal district courts adapt to this transformed legal landscape is likely to vary considerably by locale, so its impact may diverge in intensity, quality, mechanisms, and outcomes. This NIJ-funded research was designed to examine whether and in what ways these formal legal changes have trickled down into legal practice.

We conceptualize the project as comprised of two key dimensions—time and place. Our analyses track changes in sentencing outcomes (and, less directly, sentencing processes) over time in an effort to measure how these court decisions, along with other sentencing policy changes, affect sentencing behavior on the ground. Integrated into our conceptual model is the notion that place also matters—both in terms of local legal structures and norms, as well as broader socio-political cultural norms. We view the federal district court system not as a singular national legal structure with hierarchically arranged and geographically dispersed sub-units, but rather as a semi-autonomous set of systems that are loosely governed by the same formal rules and policies (this level of governance also varies over time), which are also embedded in very localized legal cultures that are themselves shaped by regionally specific historical contingencies and norms (Ulmer,

2005). Thus, we begin with the supposition that sentencing practices can potentially change not only over time in a pattern that reflects national-level impacts, but that each district and/or region may reflect qualitatively and quantitatively different reactions to macro-level forces, including the legal mandates inherent in *Booker, Gall*, and *Kimbrough*.

We limit our analyses to U.S. Sentencing Commission (USSC) federal drug trafficking case outcome data from fiscal years 1993-2009.[2] The time period was constrained by data quality and availability, as the Commission views 1992 as the first year in which the data are sufficiently complete and reliable. Nonetheless, this represents among the most comprehensive longitudinal regression studies done of the USSC sentencing outcome data to date.

We chose to limit our analysis to drug trafficking for several reasons. First, since the inception of the Guidelines, drug trafficking has been the single largest category of federally sentenced offense, so it comprises a significant share of federal court interest and resources.[3] Drug trafficking sentences grew dramatically more punitive under the Guidelines, more so than any other category of offense; federal drug policy, particularly in relation to crack cocaine, has been controversial both within the court community as well as among the general public. Drug trafficking offenses are also especially likely to be subject to mandatory minimum sentences, so provide us with a way to contextualize change since 2005, since mandatory minimums were not affected by the *Booker* line of cases. Finally, from a conceptual standpoint, there is reason to expect that the way in which drug trafficking offenses are prosecuted and adjudicated will differ from how other categories of federal offenses (such as immigration offenses and white collar economic crime) are treated, so our approach provides a cleaner test of change. As such, we take up Engen's (2011, 1141) recent suggestion that future research on the effect of *Booker*, et al. on sentencing outcomes should "consider whether the impact of *Booker* and *Gall* differs by

---

[2] We also obtained data for fiscal year 1992, but we were unable to use it because it did not including coding for mandatory minimum sentences, which is one of our key variables of interest.

[3] Immigration offenses have closed the gap on drug trafficking as an offense category, and since 2008 represent the largest category of sentenced offense, followed by drug trafficking. As such, immigration cases might have been an alternate choice, but interpreting the outcomes of immigration cases is complicated by the "fast-track" early disposition program authorized in 2003, which affects an increasing percentage of such cases. Very limited data is available on those fast-tracked cases, they are based upon an explicit agreement to give a below range sentence in exchange for deportation, and the underlying specific offense characteristics can vary considerably in those cases. In addition, the immigration cases are much more regionally variable than are the drug cases, in that a relatively small set of districts account for the bulk of such cases.

15

offense."

We recognize that formal sentence outcome data, even as complete as the USSC data are, can only imperfectly get at our more processual questions of interest. We also recognize that there are some risks in drawing conclusions from analyses that span structural changes to sentencing procedures. As Piehl and Bushway (2007) have empirically demonstrated in their work on state courts, outcome data under highly structured presumptive sentencing systems—like 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." (Piehl and Bushway, 2007, 122). In other words, comparisons of pre- and post-*Booker* outcome data for sentencing disparities without being able to account for discretionary processes that happen outside of the formal court may be misleading.

B. Literature citations and review:

Because they so dramatically changed the way sentencing is done in the federal system, a number of social scientists and legal scholars have scrutinized the Guidelines' application and impact since they were first introduced in 1987. These scholars have generally relied on the U.S. Sentencing Commission for their data, since the USSC collects detailed information from all of the federal district courts about each criminal case that reaches the sentencing stage. Among the central questions that have been addressed by this body of research are: Have sentence lengths and imprisonment risk increased under the Guidelines (Stith and Cabranes, 1998)? Has plea bargaining been reduced or reshaped by the introduction of the Guidelines (Schulhofer and Nagel, 1989; Nagel and Schulhofer, 1992)? And how have charging, charge reductions, and departures worked under the Guidelines (Johnson, Ulmer, and Kramer, 2008; Shermer and Johnson, 2009; Wilmot and Spohn, 2004). A particular interest of this research has been to assess whether and how much the Guidelines have achieved their goal of ensuring uniformity in sentencing across like offenders. Thus, a number of scholars have looked at how much and under what

conditions "extra-legal" defendant and contextual factors such as race, ethnicity, class, gender, and geography impact sentence outcomes (e.g., Albonetti, 1997; Everett and Wojtkiewicz, 2002; Kautt, 2002; Mustard, 2001; Steffensmeier & DeMuth, 2000; Wu and Spohn, 2010). In general, the findings indicate that although some forms of disparity, such as inter-judge sentencing disparities, were reduced with the implementation of the Guidelines (Anderson, Kling, and Stith 1999), race, ethnicity, and gender continue to have predictive power in terms of sentence severity, as do locale and other district-level variables (e.g. Doerner and Demuth, 2010; Feldmeyer and Ulmer, 2011; Ward, Farrell and Rousseau, 2009; Wu and Spohn, 2010).

Recent work by Johnson, Ulmer, and Kramer (2008) helps explain the persistence of disparities under the Guidelines despite the USSC's efforts to achieve uniformity. Their analysis suggests that Guidelines are sidestepped in a manner that likely serves local needs and prerogatives, and that fits with local organizational norms. It also indicates that the structure and process of plea bargain negotiations under the specific constraints of the Guidelines, and the substantial discretion held by the U.S. Attorney's offices, are key to sentence outcome variations for similar offenders (see also, Hartley, Maddan, and Spohn, 2007; Shermer and Johnson, 2009; Spohn and Fornango, 2009; Wilmot and Spohn, 2004). This work has been especially valuable in demonstrating the need to examine those processes—often hidden and less formal—that lead up to final sentencing to get a full understanding of how and why cases get resolved as they do, and why the policy goals are not fully achieved in practice.

Several scholars have supplemented the quantitative analyses with qualitative studies that provide a more textured, ideographic understanding of why the Guidelines fell short of their goals. Most notably, Ulmer (2005) used field interviews with court actors in four different federal district courts to examine how local organizational cultures shaped case outcomes. He found that despite the intent of the Guidelines to provide uniformity across disparate districts, "court community actors interpret Guidelines and other federal criminal justice policies differently, and use and transform these in a variety of ways." (Ulmer, 2005, 272). Ulmer suggests that the Guidelines helped structure the parameters of plea negotiations, but that the meanings and values assigned to the components of that structure (particularly around the items central to plea bargaining such as "substantial

assistance" departures) varied by local district.

An emerging body of empirical scholarship has also looked at the time dimension, with an eye toward how the *Booker* line of cases has changed sentencing outcomes. The Sentencing Commission researchers have conducted two analyses that aim to assess the impact of these cases on various kinds of sentencing disparities. In the year immediately following *Booker,* the Commission's analysis indicated that district courts' sentencing practices had not changed dramatically from the immediate pre-*Booker* period (Hofer, 2007; US Sentencing Commission, 2006). The Commission's most recent analysis, however, indicates an increase in sentence disparity among similar offenders (US Sentencing Commission, 2010). This most recent analysis amalgamates both offense types and prison and non-prison sentences, thus likely distorting some of ways that disparities played out,[4] thereby prompting replications that correct for some of these methodological choices.

Jeffery Ulmer and his colleagues (Ulmer, Light and Kramer, 2011a & b) have conducted the most extensive analyses in that regard. The researchers included additional relevant controls in their models, including district level controls; they also disaggregated the prison sentence length from the prison/no prison decision and lengthened the period under study to more fully capture mandatory Guidelines' sentencing. Generally, their findings indicate that racial disparities post-*Booker* and *Kimbrough/Gall* are equal or less than those observed for the period prior to the 2003 PROTECT Act,[5] with the one exception that some non-Whites' *rate* of incarceration increased after *Gall* relative to Whites (Ulmer, et al., 2011a; 2011b). They found that unwarranted sentencing disparities in the post-*Booker* period continue to be, in large part, a product of prosecutorial behavior, and that any changes brought about by the legal rulings "defy easy characterization into a convenient narrative" (Ulmer and Light, 2011, 339). Thus, they suggest that while sentence

---

[4] One major question has to do with whether "fast-track" immigration cases were excluded, thereby leaving in only the more serious or challenging immigration cases; another would ask about why at least offense categories were not disaggregated, since drug offenses, immigration and white collar offenses, to name just a few major categories, likely diverge in how they affect different demographic groups. See Ulmer & Light (2011), whose analysis of these data suggest that the *Booker/Gall* changes do not account for the disparities described by the USSC.

[5] The Feeney Amendment of the 2003 PROTECT Act further restricted judges' ability to depart downward from the Guidelines.

variations between districts, and disparities between similarly situated offenders at the individual level persist, neither has been exacerbated by *Booker*'s or *Gall*'s mandates.

Scott (2009) also empirically examined the impacts of *Booker* and *Kimbrough/Gall* on sentencing disparity, as well as the impact of the 2003 PROTECT Act which aimed to further restrict judges' ability to depart downward from the Guidelines, which preceded *Booker*. He found that the cases did not incite a sentencing revolution, but rather resulted in a slow but steady pattern of departure over time. He looked specifically to the district of Massachusetts to conduct a case study of inter-judge sentencing disparity as a function of the policy interventions (the PROTECT Act, and the *Booker* and *Gall* decisions), and found a steady increase in inter-judge disparity over time, indicating that legal change is occurring but its manifestation is more gradual than was predicted at the time of the *Booker* decision (see also, Hofer, 2007; 2011).[6] In terms of between-district variations, Farrell and Ward's (2011) work suggests that analyses attempting to isolate the impact of formal legal change are complicated by co-occurring changes to the federal system. Most notably, the demography of federal courtroom workgroups have become less diverse over the post-*Booker* period, which has been shown to impact rates of racial disparity in sentence outcomes (Ward, et al., 2009).

Using a very different methodology, Tiede (2009) compared a single case type with identical relevant sentencing factors (both in terms of offense characteristics and criminal histories) over time and across circuits. She found that the relative proportion of decisions to depart varied significantly as a function of jurisdictional geography (in this case, Circuit) both before and after *Booker*. She did not find between-circuit differences for those sentenced with either category (within range or among the departure cases) so viewed that case-level decision to depart as the significant driver of geographic disparities over time.

There is another feature of the changing landscape of federal sentencing that has been less fully integrated into this emerging body of "post-*Booker*" research. That is, as

---

[6] A significant drawback to this study is that Massachusetts is among the top 5 districts in terms of the rate of downward departures/below range sentences in the post-*Kimbrough/Gall* era across all offense types. This status is even more pronounced in drug cases: as noted below, 2008 US Sentencing Commission data indicate that 39% of all sentenced drug cases in this district were below range or departed downward for reasons other than "substantial assistance" or other government endorsement, the highest rate of all districts.

suggested earlier, that the mandatory minimums enacted by Congress in the 1980s and 1990s are left untouched by the legal decisions. In light of this complication, Fischman and Schanzenbach (2011) tested where observed racial disparities in sentence outcomes are coming from, over time, by specifying various discretionary decisions (including the application of mandatory minimums, departure decisions, safety valve use) as dependent variables. Their analyses, which only include serious violence, drugs, and weapons cases in the dataset, strongly suggest that any increased post-*Booker* sentence disparity between Blacks and Whites is due to prosecutors' expanded use of mandatory minimum statutes. Conversely, their findings on the impact of *Booker* and *Gall* "suggest that judicial discretion does not contribute to, and may in fact mitigate, racial disparities in Guidelines sentencing." (Fischman and Schanzenbach, 2011, 19). Indeed, it may be that U.S. Attorneys elect to seek mandatory minimums in more eligible cases in the post-*Booker* era as a way to avoid the uncertainties of judicial sentencing under the new "advisory" Guidelines system.

C. Statement of rationale for the research:

This project was designed to look at federal sentencing holistically as the realities of the *Booker, Gall,* and *Kimbrough* decisions trickle into daily federal court practice. We conduct quantitative analyses of U.S. Sentencing Commission individual sentence outcome data, supplemented with district level and state level variables, to assess sentence outcome variation nationally, by district, for 5 categories of drug trafficking offense types (crack cocaine, powder cocaine, heroin, methamphetamine, and marijuana)[7] subject to section 2D1 of the Guidelines, for the years 1993-2009. Research questions about the quality, variation, and extent of sentencing change over time were tested using several different analytic methods, described below, in order to examine the intersection of time and place in the federal sentencing of drug trafficking. We aimed to answer several specific research questions:[8]

---

[7] We attempted to identify MDMA cases separately, but it was not included as a primary drug, and identifying MDMA cases from other variables on drug type yielded too few cases for analysis. Therefore, we analyzed five drug categories instead of the initial six categories we had initially proposed, with MDMA considered an "other" drug.

[8] We additionally sought to answer the questions "Do districts with larger ratios of pre-Guidelines active judges demonstrate more post-*Booker/Gall/Kimbrough* variations than those with lower ratios?" and "are

1 Have federal sentences in drug trafficking cases increasingly varied from the presumptive Guidelines sentence recommendations from the pre-*Booker* period through to the post-*Kimbrough/Gall* period?

2. If so, are the increases significantly greater for cases that are not subject to mandatory minimums than those that are?

3. Do districts that handle large numbers of drug cases demonstrate less variation over time compared to districts that handle relatively fewer drug cases, as might be expected by theories about organizational norms and resistance to change? Similarly, do smaller districts demonstrate more variation across all time periods than larger districts, reflecting both a longer-standing, more individualized approach and a decreased reliance on expectations about "going rates" for prototypical cases?

4. What are the qualities of the variations: are they generally increasing or decreasing sentence outcomes for similar offenders over time?

II. Methods:

A. Data sources & variables of interest:

We obtained the Defendants Sentenced Under the Sentencing Reform Act data from the US Sentencing Commission from fiscal years 1993-2009, which contains all offenders sentenced in federal court in those fiscal years. These data include information on individual offenders, such as demographics, offense, case characteristics, and sentence. Drug trafficking cases were identified by using the sentencing guideline offense, which provided us with a sample of N=376,637 cases.[9] Because we expected variation at the federal district and state levels, we merged these data with federal district-level indicators for the 89 federal districts[10] from the Federal Court Management Statistics website, and

---

districts with the shortest aggregate mean experience of the judiciary more likely to vary due to less entrenchment within the Guidelines framework?" We were unable to identify judges or locate reliable data on length of experience on the judiciary for the entire period under study, and so excluded these research questions from our final analysis.

[9] This is the same as USSC's method of identifying drug trafficking cases using 7sentencing Guideline offenses (gdlinehi): 2D1.1, 2D1.2, 2D1.5, 2D1.6, 2D1.8, 2D1.10, 2D1.14

[10] We excluded Guam, the Mariana Islands, the US Virgin Islands, Puerto Rico, and the District of Columbia from the analysis.

state level demographic data from the U.S. Census Bureau. We conducted analyses on both the sample of individual offenders using multilevel models, as well as on just the 89 districts over time using a time series cross-sectional (TSCS) design.

Because our goal is to examine changes in the legal process, we explored three dependent variable sentencing outcomes. We first examined final prison *sentence length* in months, transformed by taking the natural log to reduce skew.[11] Because we were interested in observing how sentencing may have changed with respect to adherence to the Guidelines, we also modeled the *sentence difference* between the guideline minimum sentence and the actual sentence. Because the difference between the Guideline minimum sentence and actual sentence had a very large number of 0 values (meaning that the actual sentences were at the Guideline minimum) but ranged from -470 to 470, we recoded this difference as an ordinal variable with 13 categories: no difference, 0.1-6 months greater, 6.1-12 months greater, 12.1-24 months greater, 24.1-60 months greater, 60.1 months-120 months greater, and more than 120 months of a greater guideline minimum sentence than actual sentence. Identical categories were created for the negative values, or cases that were sentenced above the minimum guideline: 0.1-6 months less, 6.1-12 months less, 12.1-24 months less, 24.1-60 months less, 60.1 months-120 months less, and more than 120 months of a lesser guideline minimum sentence than actual sentence.[12] In other words, positive values represented defendants sentenced for shorter periods of time than the guideline minimum sentence, and negative values represented defendants sentenced for longer periods of time than the guideline minimum sentence.

Finally, to evaluate the relationship of legal mandates on mandatory minimums, we looked at the use of *mandatory minimums* as a dichotomous variable outcome for all drug trafficking cases. We coded cases with mandatory minimums as 1 if they had a drug

---

[11] As per USSC's method, we recoded life sentences and greater to 470 months. Additionally, we explored prison sentence including probation, which was coded as "0" for prison time, but the results were virtually identical. Drug trafficking cases that resulted in probation alone were less than 5% of all sentences.

[12] We recoded the *sentence difference* variable after attempting to model the sentence difference in months alone. Nearly all models for this sentence would not converge due to extremely high kurtosis (approximately 19.7). Therefore, we transformed our results into a categorical variable (with kurtosis=2.4), and modeled it as a random effects model with a continuous outcome similar to sentence length. Although treating the *sentence difference* categories as a continuous variable rather than an ordered categorical variable is not ideal, scholars have noted that with many categories (as few as 6 or 7), treating the data as ordinal vs. categorical does not have substantial difference in the outcome (Bollen & Barb, 1981; Rhemtulla, Brosseau-Liard, and Savalei, forthcoming).

mandatory minimum, and 0 otherwise. For the TSCS district analysis over time, we examined the *mean logged sentence length, mean sentence difference* and *proportion of mandatory minimums* in a district. Importantly, 64% of drug trafficking cases in these years were sentenced under mandatory minimum sentences, and so even if judges wanted to depart in these cases, they would not be authorized to do so under *Booker* or *Gall/Kimbrough*.

Our independent variables of interest are the two time periods following the *Booker* case, on January 11, 2005, and *Gall* and *Kimbrough* cases, on December 9, 2007. We also included time periods for pre-*Koon* (before June 1996), *Koon* (from June 1996-February 2003), and the PROTECT Act (March 2003-December 2004), since each of these periods represent distinct policy mandates related to federal sentencing (see Ulmer, et al, 2011a for a fuller discussion of this periodization). These variables were each coded as a dummy variable with a value of 1 for cases that occurred during each of these time periods, and 0 otherwise. Approximately 19% of the cases occurred during the *Booker,* and an additional 12% occurred during the *Gall/Kimbrough* time period. Because of our interest in the *Booker* and *Kimbrough* time periods, we excluded the PROTECT period as a reference category. For the district analysis over time in years, we specified *Booker* as 1 for the years 2005-2007 and 0 otherwise, and *Kimbrough* as 1 for 2008 and later, and 0 otherwise. The other time periods were also dummy variables, where the pre-*Koon* period was defined as before 1996, *Koon* was from 1996-2002, and the PROTECT Act period was from 2003-2004.

We also included a number of individual case characteristics in the multilevel models. Because sentences vary fairly substantially by drug, we controlled for primary drug of crack cocaine, powder cocaine, heroin, marijuana, methamphetamine, and other drugs. About one-quarter of cases were powder cocaine, and over 20% were crack cocaine. Marijuana cases represented over one-quarter of trafficking cases as well. These were modeled as a series of dummy variables, with crack cocaine excluded as a reference category. We calculated the proportion of each primary drug type by district for the TSCS district analysis.

The recommended Guidelines sentence is measured as the guideline minimum sentence in months, which we included for the sentence length and mandatory minimum

models. Because the guideline minimum sentence was used in the calculation of the size of departures, we did not include it as a predictor in this model. We included the offender's criminal history category, which ranges from 1-6, and accounts for severity, recency and number of prior convictions (USSC, 2009). We grand mean centered both variables to estimate meaningful intercepts. To test the effects of other sentencing adjustments explicitly in the sentence length models, we utilized dummy variables for mandatory minimum drug sentences, as well as safety valve application, which allow low-level, first time drug offenders relief from mandatory minimums. The safety valve application was used in 30% of cases, which included cases that directly applied an old and new safety valve (which changed in 1995), as well as cases that were coded by the USSC as fitting the old safety valve criteria.[13] Substantial assistance, and other downward departures were also similarly applied as dummy variables.[14] Over one-fourth of drug trafficking cases in the sample had substantial assistance departures, and nearly 9% had other downward departures. In theory, some cases could have both substantial assistance and other types of departures, although this occurred in very few cases.

Additionally, we also included offenders' demographic information. Gender was coded as male = 0 and female = 1, age was represented in years and centered, and education was a dummy variable with 0 as less than a high school degree, and 1 as a high school degree or greater. We coded race as a series of dummy variables for Black/African-American, Latino/Hispanic, other race, and White as a reference group. The Latino/Hispanic group represented nearly 40% of drug trafficking cases, and Black/African-American represented just over 30%. Finally, we also controlled for number of children (centered) and U.S. citizenship, with 0 as a U.S. citizen, and 1 as a non-U.S. citizen.

We expected districts to respond differently based on local legal norms and case composition and volume. Two independent variables at the district level were included in

---

[13] The old safety valve was initially an exception from mandatory minimums for first time, low-level drug offenders, but then was also expanded to also give offenders a 2-level decrease on the Guidelines offense axis as well.

[14] Although the defendant's criminal history category, mandatory minimums, and departures are used to make the presumptive sentence recommendation, we were interested in additional effects above and beyond this. After calculating a variance inflation factor, we found that collinearity was not an issue between these variables for the multilevel models. We did find high levels of correlation in the TSCS models, however, and so we removed these variables from the analysis.

our models. The criminal caseload size, adapted from Ulmer, Eisenstein and Johnson (2010), is measured by the number of criminal case filings per judgeship, and the district's relative efficiency was captured by the median time to disposition in months. The mean number of filings per judgeship ranged considerably for our sample, with a mean of just under 130, and a standard deviation of 120. To get an idea of case composition and volume, we also included trafficking cases as a percent of overall caseload, percent crack cases, and percent Black defendants. Similarly, we included drug treatment admission rate in thousands from the Treatment Episode Data Set, and the violent crime rate in thousands from the Uniform Crime Reports at the state level. These two variables were also included in our district-level models.

Finally, we expected demographic contextual effects on sentencing practices, so we calculated state-level demographic variables from the Census Bureau, such as percent in poverty, percent high school graduate, density of people (per square mile), and percent Black population. We included these in both our multilevel and TSCS analyses, and we centered all district- and state-level predictors for the multilevel models. For the TSCS models, we also wanted to look more specifically at the effects of regions, so we included dummy variables for census regions--Midwest, Northeast, and West. We excluded South as a reference category.

B. Analytic strategy:

We conducted both multilevel analyses on individual cases, as well as TSCS analyses on districts to examine trends over time and place. The first technique was to calculate the intra-class correlation (ICC) to look at 1) the variation in sentence length, difference between the guideline minimum and actual sentence length, and use of mandatory minimums in individual cases versus between districts and between states, and 2) the variation between mean sentence length, mean difference between the guideline minimum and actual sentence length, and proportion of mandatory minimums over time within districts and states versus the variation between districts and states. To do this, we ran an unconditional multilevel model of individual cases nested in districts, nested in states, and then a second model of mean logged sentence length, mean difference in sentence, and proportion of mandatory minimums per year in a district, nested in states. To

25

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

address our research questions on variation of sentencing practices post-*Booker* and *Kimbrough/Gall*, as well as cases with and without mandatory minimums, high and low proportions of drug cases, and high and low volumes of cases (measured as cases per judgeship in a district), we then divided the sample by each of these criteria and ran a series of empty models at the case and district level.[15] This enabled us to compare variation in sentencing practices between groups based on case or district characteristics.[16] Groups for high, medium, and low proportions of drug cases and volumes of cases were split according to one standard deviation or greater above, one standard deviation within, and one standard deviation or lower than the mean.[17] These analyses allowed us to directly address research questions 1-3, as delineated in the previous section of this report, in a manner not possible by the other two sets of analyses.

Calculating the ICC for each one of the sentence length, sentence difference, and mandatory minimum outcomes gave us the proportions of variation at the district-level and the state level (Rabe-Hesketh and Skrondal, 2005, p. 438): $\rho_{\text{level 2}} = \psi^{(2)}/(\psi^{(3)} + \psi^{(2)} + \theta)$ and $\rho_{\text{level 3}} = \psi^{(3)}/(\psi^{(3)} + \psi^{(2)} + \theta)$ are the equations for the logged sentence length outcome, where ψ_2 is the variance of the level-3 macro units (states), ψ_2 is the variance of the level-2 macro units (districts), and θ is the variance of the individual cases. Similarly, the conditional intraclass correlation for the dichotomous outcome of mandatory minimums in individual cases is: $\rho_{\text{level 2}} = \psi^{(2)}/(\psi^{(3)} + \psi^{(2)} + \pi^2/3)$ and $\rho_{\text{level 3}} = \psi^{(3)}/(\psi^{(3)} + \psi^{(2)} + \pi^2/3)$, where $\psi^{(3)} + \psi^{(2)} + \pi^2/3$ is the total residual variance.

We then ran a series of random effects models, with individual cases (sentences) nested in districts, nested in states.[18] In addition to accounting for variation between and

---

[15] We excluded the state level for these models examining variance over time periods, with and without mandatory minimums, proportions of drug cases, and volumes of cases because the initial empty models for the whole sample indicated very little variance at the state level. By dividing the sample down further, many of the three level models were unable to converge. Additionally, we only ran these models and calculated ICCs on the individual case outcomes of sentence length, difference in sentence length, and mandatory minimum, but not the aggregated outcomes at the district level.

[16] For example, we compared variation in sentence lengths in multiple time periods--pre-*Koon, Koon*, PROTECT Act, *Booker* and *Kimbrough/Gall—*as a function of caseload size.

[17] Because cases per judgeship in a district was highly skewed, we created the three groups by taking the natural log of the case per judgeship variable and then estimating the mean and one standard deviation above and below the mean of the logged variable.

[18] We did 3-level models for the total sentence length outcome only. While we were able to estimate the empty 3-level model for mandatory minimums, the 3-level models with predictors did not converge. While conceptually, the 3-level model is useful, it is likely that it did not converge because there is such little

within districts and over time, these models allowed us to include predictors at the individual case, district, and state levels. The general equation for the random effects model is (Rabe-Hesketh and Skrondal, 2005): $Y_{ij} = \beta_1 + \zeta_{1j} + \beta_2 X_{ij} + \zeta_{2j} X_{ij} + \varepsilon_{ij}$ for i cases in j districts, where ζ_1 is a random intercept for each district, and ζ_2 is the random slope for each j district. We included random slopes for *Booker* and *Gall/Kimbrough* time period with unstructured covariance, because we expected that the two cases would have differential impacts on sentence lengths and use of mandatory minimums.

Finally, we collapsed the individual case data into district-level data and conducted time series, cross-sectional models (TSCS) with panel-corrected standard errors.[19] TSCS models are "repeated observations on a series of fixed units" (Beck, 2001: 111).[20] In this case, the districts are the fixed units, and the repeated observations are the districts' yearly caseload characteristics. TSCS models allowed us to focus on more macro level, district-level trends over time in years, rather than focusing on individual case effects. In other words, while we lose individual case characteristic information in predicting how any one particular case is decided, the advantage of using a TSCS design is that we can focus on district-level changes over time, which is of central interest to this project. Changes in sentencing practices we observe due to *Booker* and *Kimbrough*, for example, would not likely manifest in any particular sentenced case (at least this would be impossible to observe in the individual case). On the other hand, because we aggregate the data to the district level, it makes slight changes in sentencing practices more difficult to observe than in the multilevel models. Thus, these analyses provide an additional window into how sentencing changes over time and across place, but do not fully replace the multi-level models.

---

variation at the state level compared to the district and individual case level. Instead, we estimated 2-level models for the mandatory minimum outcome.

[19] Under this model, we use OLS parameter estimates, but replace the standard errors with "panel corrected" standard errors, which are robust if disturbances are heteroskedastic and contemporaneously correlated across panels (Beck & Katz, 1995).

[20] These data are very similar (and often are used interchangeably) with panel data, where there are repeated observations on units over time. Beck (2001) notes that TSCS data often have a relatively larger number of time points and smaller sample size, and panel data often have a larger sample size and smaller number of time points. On the other hand, in panel data "units are sampled and they are typically only observed a few times. TSCS units are fixed; there is no sampling scheme for the units and any 'resampling' experiment must keep the units fixed and only resample complete units" (Beck, 2001; 113). Therefore, we stick with the TSCS characterization of the data, and in any case, there is a similar range of methods for both data.

Using the TSCS models, we examined the mean logged sentence length, the mean difference in the guideline minimum and actual sentence length, and proportion of mandatory minimum trafficking cases in districts over time. The general equation for TSCS models is (Worrall and Pratt, 2004): $y_{it} = x_{it}\beta + e_{it}$ where observations are indexed by district i and time t. We estimate a series of models for all three outcomes, the *mean logged sentence length* in districts per year, the *mean sentence difference,* and the *proportion of mandatory minimums* in districts per year.

Results:

A. Statement of results:

1) Variance at the individual case, district, and state level, and districts over time as measured by intra-class correlation (ICC)

As illustrated in Table 1, the mean sentence length for all drug trafficking cases in our sample is 81.48 months; under 1 percent of drug trafficking cases resulted in a life sentence or greater. As reflected in Table 2, when comparing logged sentence length for individual cases within districts versus between districts, only 6.6% of the variance is between districts, and 4% is between states. On the other hand, when comparing the mean logged sentence length of districts over the years, 65% of the variance is between districts or states. In other words, while there is still more variation in sentence length between individual cases than districts and states generally, there is far greater variation in mean sentence length between districts (and states) than within districts over time.

This general pattern holds for the sentence difference (the difference between the guideline minimum and actual sentence) and the use of mandatory minimums. For individual case outcomes, the small ICC for both sentence difference and mandatory minimums suggest that the variance occurs predominantly at the individual case level compared to the district and state. When examining the ICC for mean sentence difference and the proportion of mandatory minimums in districts over time, however, the proportion of variance at the district and state are significantly larger. The estimated combined variance at the district and state level for mean sentence difference is about 54%, and proportion mandatory minimum is 56%, respectively. This suggests that there are considerable between-district differences in legal outcomes, as well as relative stability of

28

outcomes rather than radical change over time (see Figure 1 below of mean sentences by drug type over time).

To examine the variation of sentencing practices for the time periods, including pre-*Koon*, *Koon*, PROTECT Act, *Booker,* and *Kimbrough/Gall*, we ran empty models on sentence length, sentence difference, and mandatory minimums, and compared the ICCs. These results are presented in Table 3. The variation in sentence length and in mandatory minimums between districts remains relatively stable over time, ranging from 11% to 13.9% for sentence length and 12% to 15.4% for mandatory minimums. While variation in the sentence difference intercept estimate is stable across all time periods, the proportion of variance between districts increases slightly over the time periods, from approximately 6% in the pre-*Koon* period to about 10% during both the PROTECT Act and *Booker* periods and decreasing to 9.5% during the *Kimbrough/Gall* period. On the other hand, variation still predominantly occurs at the individual case level for all outcomes in all time periods.

We next examined variation in sentence length and sentence difference in cases with and without mandatory minimums, presented in Table 4. For sentence length and sentence difference, there is no clear difference in the proportion of variance explained between cases that were subject to drug mandatory minimums and cases that were not sentenced according to mandatory minimums. Approximately 9-10% of variation is explained between districts for sentence length, and 6-7% of variation is explained between districts for sentence difference. In other words, variation in sentence length and sentence difference is at the individual case level, regardless of whether the cases were subject to mandatory minimum sentences or not.

Table 5 illustrates high, medium, and low proportions of drug trafficking districts, and Table 6 reflects high, medium and low volumes of cases in districts. Districts that have high relative numbers of trafficking cases and/or a high volume of cases have relatively higher proportions of variance explained in sentence length and sentence difference between districts compared to smaller districts. For example, districts with high volume of cases have 33.6% of the variance explained between districts for sentence length, compared to 10.6% in districts with medium case volume, and 3.9% in low case volume districts. It is worth noting that one exception to this pattern for sentence length in that low proportion trafficking districts, where 14.3% of the variance is between districts. In general, however,

29

higher-trafficking and higher-volume districts appear to have relatively more uniform sentencing practices than smaller districts. This greater uniformity in sentencing practices is especially pronounced in the use of mandatory minimums. For high trafficking case districts, 21% of the variance in mandatory minimums is explained between districts-- compared to about 10% for medium and low drug trafficking districts, and for high case volume districts, 26% of the variance is explained between districts, compared to 11.6% for medium volume case districts, and 9.0% for low volume districts.

2) HLM models

Table 7 shows the results for the analyses that examine predictors of total sentence length outcome. The first model includes individual-level predictors only, the second model contains both individual-level and district and state-level, and the third model also includes random slopes for *Booker* and *Kimbrough*/*Gall* at the district level.

We first review our main question of interest: whether the loosening of the sentencing guidelines in *Booker* and *Kimbrough/Gall* resulted in sentence length changes. All time periods, including *Booker* and *Kimbrough/Gall*, indicate slightly lower sentences than during the PROTECT period. *Booker* and *Kimbrough/Gall* may therefore represent a return to sentencing norms, pre-PROTECT Act. In addition, cases that were sentenced post-*Kimbrough/Gall* have relatively shorter sentences compared to *Booker*.

Individual-level predictors remained relatively consistent in influence across all three models. In concurrence with past studies that examine the effect of individual characteristics on sentence length (Everett and Wojtkiewicz, 2002; Mustard, 2001; Ulmer, Eisenstein & Johnson, 2010), our models suggest that women receive lower sentences than men, and that relative to Whites, all other racial groups have significantly longer sentences. For drug trafficking sentences in our sample, Latinos have, on average, even longer sentences than Blacks. The number of children that an offender parented had a small but significant positive impact on sentence length, and non-US citizens were sentenced more harshly than citizens. Relative to crack cocaine, powder cocaine, heroin, and marijuana cases are sentenced to significantly shorter terms. Methamphetamine trafficking cases received even greater sentence lengths than crack on average (between 1998 and 2010, crack and methamphetamine were subject to the same triggers for mandatory minimums).

30

As expected, both substantial assistance and other downward departures had large negative impacts on sentence length, and being sentenced under a drug mandatory minimum led to substantially longer sentences. Mandatory minimum direct effects are net of the interaction terms, which indicate a positive effect for all time periods compared to the PROTECT Act period. The increase in sentence lengths could be related to this slight increase in the use of mandatory minimums. Finally, those who were sentenced under the safety valve provision had shorter sentences than those who were not.

The second and third models in Table 7 included district level effects. The number of filings per judgeship was very small but positive, indicating that an increased number of filings per judgeship predicted longer sentences. Percent trafficking cases had a marginally negative significant relationship, where districts that process a higher number of trafficking cases as a proportion of their overall caseload tend to sentence more leniently. After including random slopes for *Booker* and *Kimbrough/Gall*, the median time to felony disposition and percent Black defendants in the district became nonsignificant. At the state level, states with higher treatment admissions rates, higher percentages of high school graduates, and a higher violent crime rate both positively impacted sentence length, although treatment admissions rate was a very marginal effect and not consistent between models. Similar to our individual-level findings for race, we also found race effects at the state level, in that states that had a higher percent Black population also had longer sentences.

The sentence difference models, which represent a measure of deviation from the Guidelines, are presented Table 8. In general, defendants are sentenced lower than the guideline minimum sentence, and it appears that the extent of the below-Guidelines difference has steadily increased over time. The pre-*Koon* and *Koon* time periods reflect sentencing patterns that hue closer to the Guidelines compared to the PROTECT Act period, and the *Booker* and *Kimbrough/Gall* indicate more below-Guidelines variance since the PROTECT Act.

Similar to our findings in the sentence length models, women receive larger sentence discounts relative to the Guidelines than men. Compared to white defendants, Black defendants are sentenced more closely to the Guidelines. After district and state level factors are included in the models, there is no significant difference for Latinos,

31

compared to Whites. There was no effect for non-US citizens, unlike in the sentence length models.

Compared to crack cocaine, powder cocaine, heroin, and marijuana are all sentenced more closely to the guideline minimum sentence. There is no significant difference between methamphetamine and crack cocaine in terms of sentence discounts. Although total sentence length for both crack cocaine and methamphetamine remain the longest, they are also subject to the largest deviations from the guideline minimum, which makes sense given both crack cocaine and methamphetamine's harsh guideline minimum sentences (which are most frequently mandatory minimum sentences). This may indicate that cases with drug mandatory minimums are especially prone to deviations from the Guidelines over time. In other words, it appears that court actors have collectively corrected for the extremely long sentences required by the Guidelines and mandatory minimum statutes in both crack cocaine and methamphetamine cases. As expected, substantial assistance and other downward departures have a large positive effect on the difference in sentence length.

The district and state level effects suggest the relevance of contextual processes, although there are some inconsistent effects between models 2 and 3. The case volume measure of the number of filings per judgeship changes signs and becomes only marginally significant in the final model. The median time to felony disposition indicates a positive effect, in that districts with longer case processing time tend to sentence lower relative to the Guidelines. Districts with higher percentages of drug trafficking cases sentence more closely to the Guidelines, although there is no effect for the percent of crack cases, treatment admissions rate and the percent in poverty in the final model. Places with a higher population density and a higher violent crime rate also tend to be sentenced lower relative to the Guidelines, although density effects are very marginal. Finally, similar to the individual-level findings, states with larger Black populations are sentenced more closely to the Guidelines in the final model.

Table 9 reflects results from the mandatory minimum dichotomous outcome variable. Models with random slopes for *Booker* and *Kimbrough/Gall* at the district level were unable to converge, so we present results for models with just individual-level indicators, and then district- and state-level predictors with random intercepts only. In the

individual-level model, the use of mandatory minimums appears to be higher during the *Koon, Booker* and *Kimbrough/Gall* periods compared to the PROTECT Act period; the pre-*Koon* time period indicates a lowered rate of mandatory minimum use.[21] This may suggest that prosecutorial behavior continues to mediate sentence outcomes in the post-*Booker* period through the pursuit of more mandatory minimum sentences than before, or it could represent a return to the case processing norms that prevailed before the PROTECT Act, as was similarly observed in the sentence length models.

Women have far lower odds of receiving a mandatory minimum sentence compared to men. Non-US citizens have higher odds of receiving a mandatory minimum sentence for trafficking compared to citizens. Blacks have nearly even odds compared to Whites in receiving mandatory minimum sentences. This means that Blacks are being sentenced, all else equal, for longer sentences even though they are not more likely to be sentenced under mandatory minimums compared to Whites. Latinos, on the other hand, are consistently significantly more likely to receive mandatory minimum sentences than Whites. As expected, those with more serious criminal history scores had higher odds of receiving a mandatory minimum sentence.

Reflecting the harsh mandatory minimum sentencing for crack cocaine, all other drugs, including methamphetamine, have lower odds of receiving a mandatory minimum sentence. Thus, even if final sentences were longer for methamphetamine than crack cocaine on average, this was not due to greater use of mandatory minimum sentences. Conversely, crack cocaine sentences are most likely driven by mandatory minimums compared to other drugs, net of race.   Given that a disproportionate amount (approximately 85%) of prosecuted crack cocaine trafficking offenders were Black in our sample, these set of findings could explain why, after controlling for drug type, there is little Black-White difference in mandatory minimums.

The two types of downward departure in the model had consistent and strong opposite effects; cases with substantial assistance actually had higher odds of being sentenced under mandatory minimum sentences for drug trafficking, whereas other types of downward departures had lower odds of being sentenced to mandatory minimums. It is

---

[21] The lower relative use of mandatory minimums in this early period may be, in part, an artifact of the overall increase of such statutes over time. Congress has passed numerous new mandatory minimum laws since the 1980s, but it does not usually rescind such laws once on the books.

likely that cases with substantial assistance were among the most serious cases, in which the defendants faced very long sentences so were especially motivated to provide information to the government to reduce their exposure in sentencing.[22]

Model 2 indicates the district- and state-level effects for mandatory minimums. Number of filings per judgeship predicted marginal decreased odds of receiving a mandatory minimum sentence. This suggests that busier courts are slightly less likely to use mandatory minimum sentences than less busy courts, but the difference is so small (the odds ratio is .999) as to not be very meaningful. All else equal, longer median time to felony disposition predicted lower odds of receiving a mandatory minimum sentence for trafficking.

At the state level, the drug treatment admissions rate did not significantly affect mandatory minimums. Poverty and urban density had small but significant negative effects, indicating that district courts in poorer, more urban, states were less likely to sentence with mandatory minimums. There was no effect of percent Black population in the state on the relative use of mandatory minimum sentences.

Overall, there was a counterintuitive finding that the use of mandatory minimums was positively related to deviations from the Guideline minimums. This suggests that prosecutorial behavior is especially influential in sentence deviations. This was especially the case in the years before the PROTECT Act. Additionally, contrary to what might be expected, sentence deviations in the *Booker* and *Kimbrough* periods are no more likely in non-mandatory minimum Guidelines cases than in mandatory minimum cases. Thus, it does not appear the new-found judicial discretion in non-mandatory minimum cases is directly responsible for increases in sentence deviations post-*Booker*.

Measured differences from the guideline minimum sentence steadily increased over time, although the nature of those increases appear to have changed from pre-PROTECT Act to post-*Booker*. This is reflected in the multi-level models measuring the interactive effect of mandatory minimums x time period on the sentence difference variable. Specifically, in the years before the PROTECT Act was instituted, increased application of mandatory minimums was associated with greater deviations down from the guideline minimum. As noted above, this effect disappeared after *Booker*. This finding suggests that

---

[22] It appears from the data that this is the case.

the actual mechanisms for achieving desired sentence outcomes changed as a function of policy. Thus, prior to 2005, US Attorneys held the most discretionary power to seek particular sentence outcomes, so sentence discounts in both mandatory minimum cases and regular Guidelines cases were primarily authorized by prosecutors. Once some discretionary sentencing power in Guidelines cases shifted to judges in the wake of Booker, prosecutors likely use mandatory minimums more strategically to obtain particular outcomes, so are less willing to discount those sentences.

3) TSCS Models

Table 10 lays out the results from the TSCS models with panel-corrected standard errors for mean logged sentence length. In model 1, we only included time and drug effects. Models 2 and 3 include the same predictors, but model 2 assumes no autocorrelation within districts, and model 3 specifies a first-order correlation within districts.[23] Thus, the r-squared value for model 3 is significantly higher than the other two models.

With the PROTECT Act as a reference category, initial time effects for *Booker* and *Kimbrough/Gall* disappear after controlling for judicial variables, state demographics, and region. In other words, there does not appear to be any significant change in mean sentence length over time, regardless of the *Booker* and *Kimbrough* decisions. There appears to be somewhat inconsistent evidence for the time periods before the PROTECT Act; the *Koon* era indicated slightly shorter sentences compared to the PROTECT Act, but no other time periods indicated any differences. While the sentence length multilevel models indicate slightly shorter sentences both in the pre- and post-PROTECT Act periods, these differences are not detected in the more aggregated TSCS models.

---

[23] A reviewer suggested examining more than first-order AR errors, with the reasoning that the impact of the independent variables might be realized later. Unfortunately, Stata does not offer more than a first-order autocorrelation with the xtpcse command. We did model dynamic TSCS models with 1 year lags for all outcomes to examine the effects of the time periods on mean sentence length, difference in sentence length, and proportion mandatory minimum in the following year. For the time period outcomes, the results were similar, except Kimbrough/Gall indicated small but significant increases in sentence length. For sentence difference, Booker showed marginally lower effects for sentence difference in contrast to model 3, and Koon showed marginally higher effects. For the mandatory minimum models, *Kimbrough* had a positive significant effect, and pre-*Koon* had a marginally significant negative effect. Given our approach to modeling the time periods, however, rather than individual events in a single year (e.g. a "Booker" or "Kimbrough/Gall" year), these models are not clearly interpreted, as the effects of a time period could still impact some of the sentencing outcomes in the same year.

Similar to the results from the sentence length multilevel models, higher proportions of crack trafficking cases consistently predict longer mean sentence length in districts compared to all other drugs except methamphetamine. In the multilevel models, methamphetamine predicted a significantly longer sentence than crack cocaine, but the TSCS models show little to no difference between methamphetamine and crack cocaine on mean sentence length in the district. This suggests that both crack cocaine and methamphetamine drive similar mean sentence lengths at the district level, all else equal.

The number of filings per judgeship had a small but significant negative effect on mean sentence length, so in districts where caseload size was higher, mean sentence lengths were lower. In contrast, the median time to felony disposition had no consistent impact across models 2 and 3. Consistent with the multilevel models, higher percentages of high school graduates were associated with longer sentences. Higher levels of poverty were associated with lower mean sentence length, and higher violent crime rates predicted longer mean sentences for trafficking in the district models. States that had lower treatment admissions rates had higher mean sentence lengths, which could indicate that a given locale's relative commitment to and support of drug treatment shapes federal sentencing within that locale. Finally, percent Black was nonsignificant in the TSCS models. We also included dummy variables for the Midwest, Northeast, and West to examine regional effects. All regions have lower mean sentences for trafficking than the South, although sentence lengths in the Midwest are not significantly lower in the final model.

In the second set of TSCS models, we examined the mean difference in sentence length in the same manner as we did for the mean logged sentence outcome (Table 11). Compared to the PROTECT Act period, there is a weak positive effect for *Booker* and *Kimbrough/Gall* across the models. Even though mean sentence length did not change at the district level overall, defendants were being sentenced lower relative to the guideline minimum sentence over time. This effect appears to be a gradual rather than dramatic change, which is concurrent with the progressively increasing sentence length difference in the HLM models over time.

In the final models, there were no differences between crack cocaine and other drugs in how defendants were sentenced relative to the guideline minimum sentence. Thus, even given sentence disparities by drug, they seem to be driven by (or at least related to)

already-existing disparities in the Guidelines or in mandatory minimum sentences. The differences found in the HLM models between powder cocaine, heroin, and marijuana, which were sentenced more closely to the Guidelines compared to crack cocaine, disappeared in these aggregated models.

We found a positive effect for the median time to felony disposition, suggesting that districts where cases take longer to process result in lower sentences compared to the Guideline minimum sentence. All other district and state level predictors were nonsignificant in the final model, however. Similar to the sentence length models, relative to the South, all other regions indicated greater sentence reductions from the Guidelines.

Finally, we examined the proportion of mandatory minimum sentences for trafficking in districts over time. Unlike the individual-level case models, there are no significant effects over time in the final model, suggesting that while there may have been some increased use of mandatory minimums over time in some cases, there was not a large enough effect to detect it at this aggregated level. While Models 1 and 2 show small significant increases for mandatory minimum use during the *Koon*, *Booker* and *Kimbrough/Gall*, this effect does hold across the final model. Similar to the difference in sentence length, to the degree that there may be an effect of the use of mandatory minimums over time, it is likely that the PROTECT Act was a more unusual period relative to the rest of the Guidelines era, rather than a *Booker* or *Kimbrough/Gall* effect. This may support the findings in the individual-level case models that three of the four time periods had greater mandatory minimum use compared to the PROTECT Act period.

Interestingly, while the drug type results converge to some degree with the individual case models, districts with higher proportions of methamphetamine cases predict higher proportion of mandatory minimums relative to districts with higher proportions of crack cocaine. In the individual case level, crack cocaine trafficking cases have higher odds of being sentenced under a mandatory minimum compared to all other drug types. These somewhat contradictory findings at different levels of aggregation could reflect the highly regional nature of drug prosecutions at the district level. This is supported by our finding on the effect of region; judicial districts in Midwestern states had significantly higher proportion of mandatory minimums relative to the South, and the Northeast had significantly lower rates. Thus, as illustrated in Illustration 2, those districts that pursue the

37

highest percentage of crack cocaine and methamphetamine cases were more likely to use mandatory minimums.

Consistent with the multilevel models, both higher number of filings per judgeship and longer median time to disposition were related to higher rates of mandatory minimums in districts. State demographics were fairly consistent with mean sentence length, with the exception of percent Black. In this case, higher percentage Black population in the state is significantly related to a higher proportion of mandatory minimums, whereas there were no significant effects in the individual case models.

B. Tables:

Table 1: Descriptive Statistics of Dependent and Independent Variables, FY 1993-2009

Variable	Mean	SD	min	Max
Dependent variables				
Sentence length (months)	81.48	78.83	0.03	470
Difference in presumptive and actual sentence	18.61	44.56	-464	469.97
Mandatory minimum applied	0.64	0.48	0	1
Individual-level variables (N=379,808)				
Pre-Koon	0.19	0.40	0	1
Koon	0.38	0.49	0	1
PROTECT Act	0.11	0.32	0	1
Booker	0.19	0.39	0	1
Kimbrough	0.12	0.32	0	1
Female	0.13	0.33	0	1
Race				
White (reference)	0.27	0.44	0	1
Black/African-American	0.31	0.46	0	1
Latino/Hispanic	0.40	0.49	0	1
Other race	0.02	0.15	0	1
Age	33.02	9.78	16	98
HS graduate	0.51	0.50	0	1
No. dependents	1.73	2.56	0	98
US citizen	0.29	0.45	0	1
Primary drug				
Crack cocaine (reference)	0.22	0.41	0	1
Powder cocaine	0.25	0.43	0	1
Heroin	0.08	0.26	0	1
Marijuana	0.27	0.44	0	1
Methamphetamine	0.15	0.35	0	1
Other drug	0.04	0.19	0	1
Presumptive sentence	99.92	97.95	0	470
Criminal history	2.18	1.62	1	6
Substantial assistance	0.27	0.45	0	1
Other downward departure	0.09	0.28	0	1
Safety valve applied	0.30	0.46	0	1
District level variables (N=89)				
No. filings per judgeship	129.77	120.47	18	543
Median time to felony disposition	7.43	2.82	1.70	19.50
Percent drug trafficking cases	39.98	11.18	8.70	76.19
Percent crack cases	22.96	18.68	0	86.74
Percent Black defendants	26.40	17.35	0	71.64

State level variables (N=50)

Treatment admissions rate in thousands	5.72	3.86	0.48	20.04
% Poverty	13.75	3.05	4.50	26.40
% HS Grad	82.51	4.23	67.10	92.80
Violent crime rate in thousands	555.25	211.26	66.90	1207.20
Density	183.42	173.86	1.03	1173.97
% Black	12.94	7.56	0.28	37.28

Table 2: Unconditional sentence length (ln), sentence difference, and drug mandatory minimum models

	individual cases (N=376,637)		mean or proportion of cases in a year by district (N=1,602)	
	Estimate	SE	Estimate	SE
Sentence length				
Fixed effects				
Intercept: β_1	4.106	0.040	4.089	0.039
Random effects				
Level 3: $\sqrt{\Psi_3}$	0.198	0.043	0.202	0.040
Level 2: $\sqrt{\Psi_2}$	0.247	0.027	0.222	0.026
Level 1: $\sqrt{\theta}$	0.906	0.001	0.219	0.004
ρ_3	0.043		0.294	
ρ_2	0.066		0.358	
Sentence difference				
Fixed effects				
Intercept: β_1	8.310	0.077	8.291	0.074
Random effects				
Level 3: $\sqrt{\Psi_3}$	0.388	0.077	0.384	0.071
Level 2: $\sqrt{\Psi_2}$	0.467	0.050	0.421	0.048
Level 1: $\sqrt{\theta}$	2.262	0.003	0.531	0.009
ρ_3	0.027		0.243	
ρ_2	0.040		0.292	
Mandatory minimum				
Fixed effects				
Intercept: β_1	0.761	0.079	0.659	0.016
Random effects				
Level 3: $\sqrt{\Psi_3}$	0.392	0.086	0.079	0.018
Level 2: $\sqrt{\Psi_2}$	0.497	0.054	0.101	0.012
Level 1: $\sqrt{\theta}$			0.113	0.002
ρ_3	0.036		0.213	
ρ_2	0.057		0.350	

Table 3: Unconditional sentence length (ln), sentence difference and mandatory minimum models for individual case outcomes for time periods

	Pre-Koon		Koon		PROTECT		Booker		Kimbrough/Gall	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Sentence length										
Fixed effects										
Intercept: β_1	4.061	0.035	4.090	0.035	4.161	0.038	4.198	0.039	4.148	0.038
Random effects										
Level 2: $\sqrt{\Psi_2}$	0.323	0.025	0.332	0.025	0.349	0.027	0.363	0.028	0.356	0.027
Level 1: $\sqrt{\theta}$	0.914	0.002	0.881	0.002	0.878	0.003	0.903	0.002	0.916	0.003
ρ_2	0.111		0.124		0.136		0.139		0.131	
Sentence difference										
Fixed effects										
Intercept: β_1	8.153	0.067	8.286	0.067	8.130	0.078	8.380	0.081	8.540	0.077
Random effects										
Level 2: $\sqrt{\Psi_2}$	0.619	0.048	0.630	0.048	0.723	0.057	0.755	0.058	0.713	0.056
Level 1: $\sqrt{\theta}$	2.445	0.007	2.214	0.004	2.137	0.007	2.206	0.006	2.206	0.008
ρ_2	0.060		0.075		0.103		0.105		0.095	
Mandatory minimum										
Fixed effects										
Intercept: β_1	0.518	0.079	0.774	0.072	0.690	0.080	0.948	0.074	0.951	0.083
Random effects										
Level 2: $\sqrt{\Psi_2}$	0.734	0.056	0.671	0.051	0.739	0.057	0.686	0.053	0.775	0.060
ρ_2	0.141		0.120		0.142		0.125		0.154	

Table 4: Unconditional sentence length (ln) and sentence difference models for individual case outcomes for mandatory minimum cases

	MM applied		MM not applied	
	Estimate	SE	Estimate	SE
Sentence length				
Fixed effects				
Intercept: β_1	4.380	0.027	3.57	0.03
Random effects				
Level 2: $\sqrt{\Psi_2}$	0.253	0.019	0.285	0.022
Level 1: $\sqrt{\theta}$	0.796	0.001	0.841	0.002
ρ_2	0.092		0.103	
Sentence difference				
Fixed effects				
Intercept: β_1	8.564	0.074	7.704	0.050
Random effects				
Level 2: $\sqrt{\Psi_2}$	0.695	0.053	0.466	0.360
Level 1: $\sqrt{\theta}$	2.420	0.004	1.737	0.004
ρ_2	0.076		0.067	

Table 5: Unconditional sentence length (ln), sentence difference and mandatory minimum models for individual case outcomes for high, medium, and low drug trafficking districts

	High % Trafficking		Med % Trafficking		Low % Trafficking	
	Estimate	SE	Estimate	SE	Estimate	SE
Sentence length						
Fixed effects						
Intercept: β_1	4.099	0.053	4.125	0.035	4.132	0.053
Random effects						
Level 2: $\sqrt{\Psi_2}$	0.378	0.037	0.322	0.025	0.362	0.038
Level 1: $\sqrt{\theta}$	0.872	0.002	0.920	0.001	0.885	0.003
ρ_2	0.158		0.109		0.143	
Sentence difference						
Fixed effects						
Intercept: β_1	8.280	0.010	8.305	0.069	8.261	0.092
Random effects						
Level 2: $\sqrt{\Psi_2}$	0.719	0.072	0.629	0.049	0.615	0.068
Level 1: $\sqrt{\theta}$	2.277	0.005	2.260	0.003	2.161	0.008
ρ_2	0.091		0.072		0.075	
Mandatory minimum						
Fixed effects						
Intercept: β_1	0.671	0.132	0.768	0.066	0.761	0.093
Random effects						
Level 2: $\sqrt{\Psi_2}$	0.926	0.094	0.608	0.048	0.600	0.069
ρ_2	0.207		0.101		0.099	

Table 6: Unconditional sentence length (ln), sentence difference and mandatory minimum models for individual case outcomes for high, medium, and low case volume districts

	High case volume		Med case volume		Low case volume	
	Estimate	SE	Estimate	SE	Estimate	SE
Sentence length						
Fixed effects						
Intercept: β_1	3.701	0.244	4.135	0.034	4.150	0.028
Random effects						
Level 2: $\sqrt{\Psi_2}$	0.644	0.173	0.307	0.024	0.194	0.021
Level 1: $\sqrt{\theta}$	0.906	0.002	0.890	0.001	0.962	0.003
ρ_2	0.336		0.106		0.039	
Sentence difference						
Fixed effects						
Intercept: β_1	8.077	0.219	8.271	0.067	8.189	0.093
Random effects						
Level 2: $\sqrt{\Psi_2}$	0.577	0.155	0.610	0.048	0.646	0.068
Level 1: $\sqrt{\theta}$	1.759	0.005	2.344	0.003	2.502	0.008
ρ_2	0.097		0.063		0.062	
Mandatory minimum						
Fixed effects						
Intercept: β_1	0.510	0.407	0.750	0.072	0.732	0.083
Random effects						
Level 2: $\sqrt{\Psi_2}$	1.075	0.291	0.657	0.051	0.572	0.060
ρ_2	0.260		0.116		0.090	

Table 7: HLM results for total sentence length (ln) (N=376,637)

	Model 1: individual predictors			Model 2: district and state predictors			Model 3: Random slopes		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Individual level effects									
Pre-Koon	-0.077	0.006	***	-0.088	0.008	***	-0.097	0.008	***
Koon	-0.090	0.006	***	-0.086	0.006	***	-0.088	0.006	***
Booker	-0.065	0.006	***	-0.061	0.006	***	-0.067	0.012	***
Kimbrough/Gall	-0.115	0.007	***	-0.107	0.007	***	-0.129	0.015	***
Age	0.002	0.000	***	0.002	0.000	***	0.002	0.000	***
Female	-0.176	0.003	***	-0.175	0.003	***	-0.176	0.003	***
HS graduate	-0.003	0.002		-0.004	0.002		-0.004	0.002	
No. dependents	0.005	0.001	***	0.005	0.001	***	0.005	0.001	***
Non-US citizen	0.041	0.003	***	0.041	0.003	***	0.040	0.003	***
Race: Black	0.023	0.003	***	0.023	0.003	***	0.022	0.003	***
Race: Latino/Hispanic	0.052	0.003	***	0.050	0.003	***	0.048	0.003	***
Race: Other	0.035	0.007	***	0.030	0.007	***	0.027	0.007	***
Powder cocaine	-0.013	0.003	***	-0.012	0.003	***	-0.012	0.003	***
Heroin	-0.015	0.005	**	-0.017	0.005	***	-0.019	0.005	***
Marijuana	-0.302	0.004	***	-0.300	0.004	***	-0.298	0.004	***
Methamphetamine	0.091	0.004	***	0.091	0.004	***	0.089	0.004	***
Other drug	0.035	0.006	***	0.038	0.006	***	0.038	0.006	***
Presumptive sentence	0.006	0.000	***	0.006	0.000	***	0.006	0.000	***
Criminal history	0.042	0.001	***	0.042	0.001	***	0.042	0.001	***
Substantial assistance	-0.586	0.002	***	-0.587	-0.002	***	-0.589	0.002	***
Other downward departure	-0.352	0.004	***	-0.347	0.004	***	-0.349	0.004	***
Drug MM applied	0.434	0.006	***	0.434	0.006	***	0.438	0.006	***
Safety valve applied	-0.239	0.003	***	-0.237	0.003	***	-0.237	0.003	***

Pre-Koon x drug MM	0.062	0.008	***	0.066	0.008	***	0.068	0.008	***
Koon x drug MM	0.079	0.007	***	0.081	0.007	***	0.078	0.007	***
Booker x drug MM	0.038	0.008	***	0.042	0.008	***	0.033	0.008	***
Kimbrough/Gall x drug MM	0.054	0.008	***	0.053	0.008	***	0.054	0.009	***
District level effects									
No. filings per judgeship				0.000	0.000	***	0.000	0.000	***
Median time to felony disposition				-0.006	0.001	***	0.000	0.001	
Percent trafficking cases				-0.001	0.000	**	-0.001	0.000	**
Percent crack cases				0.000	0.000		0.000	0.000	
Percent Black defendants				-0.001	0.000	***	0.000	0.000	
State level effects									
Treatment admissions rate				0.001	0.001		0.003	0.001	**
% Poverty				0.003	0.001	***	0.001	0.001	
% HS Grad				0.009	0.001	***	0.007	0.001	***
Violent crime rate				0.000	0.000	***	0.000	0.000	***
Density				0.000	0.000		0.000	0.000	
% Black				0.011	0.002	***	0.011	0.002	***
Constant	4.046	0.018	***	4.097	0.020	***	4.067	0.020	***
State level variance Ψ	0.008	0.003		0.009	0.003		0.009	0.003	
District level									
Post-Booker variance Ψ_{33}							0.008	0.001	
Post-Kimbrough variance Ψ_{22}							0.013	0.002	
Constant variance Ψ_{11}	0.008	0.002		0.007	0.002		0.005	0.001	
$\rho32$							0.009	0.002	
$\rho31$							0.001	0.001	
$\rho21$							0.001	0.001	
Individual level variance θ	0.271	0.001		0.270	0.001		0.268	0.001	

***$p<0.001$; **$p<0.01$; *$p<0.05$

Table 8: HLM results for sentence difference (N=376,637)

	Model 1: individual predictors			Model 2: district and state predictors			Model 3: Random slopes		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Individual level effects									
Pre-Koon	-0.154	0.018	***	-0.098	0.022	***	-0.019	0.022	
Koon	-0.189	0.016	***	-0.173	0.017	***	-0.143	0.017	***
Booker	0.353	0.017	***	0.320	0.017	***	0.261	0.038	***
Kimbrough/Gall	0.491	0.019	***	0.479	0.020	***	0.471	0.048	***
Age	-0.001	0.000	*	-0.001	0.000	*	-0.001	0.000	*
Female	0.155	0.009	***	0.158	0.009	***	0.158	0.009	***
HS graduate	0.028	0.006	***	0.029	0.006	***	0.030	0.006	***
No. dependents	-0.005	0.001	***	-0.005	0.001	**	-0.004	0.001	**
Non-US citizen	-0.008	0.008		-0.007	0.008		-0.013	0.008	
Race: Black	-0.032	0.009	***	-0.037	0.010	***	-0.027	0.010	**
Race: Latino/Hispanic	-0.019	0.009	*	-0.009	0.009		0.006	0.009	
Race: Other	-0.009	0.020		0.002	0.020		0.000	0.020	
Powder cocaine	-0.127	0.009	***	-0.124	0.009	***	-0.122	0.009	***
Heroin	-0.227	0.013	***	-0.220	0.013	***	-0.215	0.013	***
Marijuana	-0.211	0.011	***	-0.212	0.011	***	-0.210	0.011	***
Methamphetamine	-0.015	0.012		-0.010	0.012		-0.001	0.012	
Other drug	-0.035	0.018		-0.038	0.018	*	-0.030	0.018	
Criminal history	0.045	0.002	***	0.046	0.002	***	0.045	0.002	***
Substantial assistance	3.907	0.007	***	3.908	0.007	***	3.924	0.007	***
Other downward departure	2.591	0.011	***	2.614	0.011	***	2.687	0.011	***
Drug MM applied	0.269	0.017	***	0.276	0.017	***	0.248	0.017	***
Safety valve applied	0.008	0.007		-0.002	0.007		-0.014	0.007	
Pre-Koon x drug MM	0.282	0.022	***	0.279	0.022	***	0.294	0.022	***

	Est.	SE		Est.	SE		Est.	SE	
Koon x drug MM	0.189	0.019	***	0.174	0.019	***	0.200	0.019	***
Booker x drug MM	-0.064	0.021	**	-0.064	0.021	**	-0.002	0.022	
Kimbrough/Gall x drug MM	-0.042	0.023		-0.051	0.023	*	-0.030	0.024	
District level effects									
No. filings per judgeship				-0.001	0.000	***	0.000	0.000	*
Median time to felony disposition				0.026	0.002	***	0.023	0.003	***
Percent trafficking cases				-0.002	0.000	***	-0.004	0.001	***
Percent crack cases				-0.002	0.000	***	-0.001	0.000	
Percent Black defendants				0.011	0.001	***	0.004	0.001	***
State level effects									
Treatment admissions rate				0.010	0.002	***	-0.001	0.003	
% Poverty				-0.014	0.002	***	-0.002	0.003	
% HS Grad				-0.006	0.002	**	0.015	0.002	***
Violent crime rate				0.000	0.000	***	0.000	0.000	***
Density				0.000	0.000	*	0.000	0.000	*
% Black				-0.009	0.005		-0.010	0.004	**
Constant	6.702	0.047	***	6.409	0.061	***	6.624	0.048	***
State level variance Ψ	0.037	0.019		0.076	0.036		0.010	0.013	
District level									
Post-Booker variance Ψ_{33}							0.092	0.015	
Post-Kimbrough variance Ψ_{22}							0.156	0.026	
Constant variance Ψ_{11}	0.095	0.020		0.096	0.022		0.089	0.019	
$\rho 32$							0.092	0.017	
$\rho 31$							0.015	0.012	
$\rho 21$							-0.021	0.016	
Individual level variance θ	2.096	0.005		2.089	0.006		2.064	0.005	

***$p<0.001$; **$p<0.01$; *$p<0.05$

Table 9: HLM results for mandatory minimum sentence (N=376,637)

	Model 1: individual predictors				Model 2: district and state predictors			
	OR	Beta	SE	P	OR	Beta	SE	P
Individual level effects								
Pre-Koon	0.759	-0.276	0.018	***	0.762	-0.272	0.029	***
Koon	1.177	0.163	0.015	***	1.167	0.154	0.018	***
Booker	1.260	0.231	0.017	***	1.240	0.215	0.018	***
Kimbrough/Gall	1.363	0.310	0.019	***	1.448	0.370	0.022	***
Age	0.999	-0.001	0.000	**	0.998	-0.002	0.001	**
Female	0.762	-0.272	0.014	***	0.763	-0.271	0.014	***
Race: Black	0.955	-0.046	0.016	**	0.956	-0.045	0.017	**
Race: Latino/Hispanic	1.242	0.217	0.015	***	1.242	0.216	0.015	***
Race: Other	0.919	-0.085	0.033	*	0.909	-0.096	0.034	**
HS graduate	1.003	0.003	0.010		1.003	0.003	0.010	
No. dependents	1.009	0.009	0.003	***	1.010	0.010	0.003	***
Non-US citizen	1.140	0.131	0.013	***	1.147	0.137	0.013	***
Powder cocaine	0.803	-0.219	0.017	***	0.798	-0.225	0.018	***
Heroin	0.678	-0.388	0.023	***	0.675	-0.393	0.024	***
Marijuana	0.401	-0.913	0.019	***	0.400	-0.915	0.019	***
Methamphetamine	0.630	-0.462	0.022	***	0.627	-0.467	0.022	***
Other drug	0.031	-3.459	0.035	***	0.031	-3.477	0.035	***
Presumptive sentence	1.031	0.030	0.000	***	1.031	0.030	0.000	***
Criminal history	0.584	-0.538	0.004	***	0.582	-0.540	-0.004	***
Substantial assistance	1.657	0.505	0.012	***	1.647	0.499	0.012	***
Other downward departure	0.712	-0.340	0.017	***	0.747	-0.292	0.018	***

District level effects				1.000			
No. filings per judgeship				0.999	-0.001	0.000	***
Median time to felony disposition				1.028	0.028	0.005	***
Percent trafficking cases				0.990	-0.010	0.001	***
Percent crack cases				1.001	0.001	0.001	
Percent Black defendants				0.995	-0.005	0.001	***
State level effects				1.000			
Treatment admissions rate				1.004	0.004	0.005	
% Poverty				0.968	-0.032	0.004	***
% HS Grad				0.984	-0.016	0.004	***
Violent crime rate				1.000	0.000	0.000	
Density				0.996	-0.004	0.001	***
% Black				0.988	-0.013	0.010	
Constant	4.669	1.541	0.066	7.086	1.958	0.120	***
District level variance Ψ_{11}	1.404	0.339	0.051	2.892	1.062	0.374	

***$p<0.001$; **$p<0.01$; *$p<0.05$

Table 10: TSCS Models for mean sentence length (ln) (N=1602)

	Model 1: time & drug			Model 2: district and state			Model 3: AR(1)		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Pre-Koon	0.046	0.021	*	0.011	0.029		-0.007	0.048	
Koon	-0.046	0.017	**	-0.07	0.026	**	-0.049	0.039	
Booker	0.010	0.019		0.023	0.026		0.004	0.039	
Kimbrough/Gall	-0.036	0.02		0.014	0.028		-0.017	0.048	
Proportion Cocaine	-0.444	0.082	***	-0.395	0.042	***	-0.301	0.052	***
Proportion Heroin	-1.251	0.106	***	-0.644	0.070	***	-0.599	0.086	***
Proportion Marijuana	-1.673	0.069	***	-1.226	0.087	***	-0.826	0.081	***
Proportion Meth	-0.172	0.048	***	0.136	0.061	*	0.052	0.069	
Proportion other drug	-1.252	0.18	***	-1.046	0.125	***	-0.679	0.129	***
No. filings per judgeship				-0.001	0.000	***	-0.001	0.000	***
Median time to felony disposition				-0.011	0.002	***	0.000	0.004	
% HS Grad				0.007	0.003	**	0.007	0.003	**
% Poverty				-0.011	0.002	***	-0.009	0.002	***
Violent crime rate				0.000	0.000	***	0.000	0.000	***
Treatment admissions rate				-0.012	0.001	***	-0.007	0.002	**
% Black				0.000	0.001		0.002	0.001	
Density				0.000	0.000	***	0.000	0.000	***
Midwest				-0.060	0.009	***	-0.047	0.025	
Northeast				-0.302	0.033	***	-0.358	0.028	***
West				-0.250	0.012	***	-0.235	0.023	***
Constant	4.701	0.036	**	4.171	0.255	***	3.963	0.241	***
	827.3			217541			18790		
Chi-square	3			.03			0.84		
R2	0.444			0.599			0.860		

***p<0.001; **p<0.01; *p<0.05

Table 11: TSCS Models for mean sentence difference (N=1602)

	Model 1: time and drug			Model 2: district and state			Model 3: AR(1)		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Pre-Koon	-0.002	0.061		0.188	0.068	**	0.120	0.095	
Koon	0.163	0.056	**	0.226	0.064	***	0.131	0.073	
Booker	0.223	0.063	***	0.138	0.070	*	0.157	0.072	*
Kimbrough/Gall	0.373	0.069	***	0.315	0.075	***	0.241	0.090	**
Proportion Cocaine	0.153	0.153		-0.107	0.095		0.159	0.085	
Proportion Heroin	1.584	0.249	***	0.092	0.105		0.005	0.133	
Proportion Marijuana	0.212	0.087	*	0.222	0.149		-0.062	0.133	
Proportion Meth	0.075	0.090		-0.213	0.095	*	-0.096	0.090	
Proportion Other drug	-0.236	0.365		-0.278	0.320		-0.236	0.213	
No. filings per judgeship				0.001	0.000	***	0.001	0.000	
Median time to felony disposition				0.091	0.005	***	0.073	0.007	***
% HS Grad				-0.003	0.006		-0.001	0.007	
% Poverty				-0.027	0.008	**	-0.012	0.007	
Violent crime rate				0.000	0.000		0.000	0.000	
Treatment admissions rate				0.004	0.003		0.003	0.004	
% Black				0.006	0.003	*	0.006	0.005	
Density				0.000	0.000		0.000	0.000	
Midwest				0.119	0.024	***	0.163	0.061	**
Northeast				0.238	0.029	***	0.290	0.051	***
West				0.329	0.020	***	0.372	0.072	***
Constant	7.236	0.091	***	7.001	0.653	***	6.844	0.692	***
Chi-square	139.42			345596			6618.99		
R2	0.098			0.3			0.86		

***p<0.001; **p<0.01; *p<0.05

Table 12: TSCS Models for proportion mandatory minimum (N=1602)

	Model 1: time and drug			Model 2: no autocorrelation			Model 3: AR(1)		
	Beta	SE	P	Beta	SE	P	Beta	SE	P
Pre-Koon	-0.016	0.015		0.002	0.008		-0.011	0.020	
Koon	0.027	0.013	*	0.032	0.007	***	0.012	0.014	
Booker	0.037	0.015	*	0.037	0.005	***	0.023	0.015	
Kimbrough/Gall	0.045	0.016	**	0.065	0.005	***	0.031	0.017	
Proportion Cocaine	0.092	0.044	*	0.049	0.034		-0.022	0.031	
Proportion Heroin	-0.101	0.039	**	-0.104	0.034	**	-0.131	0.056	*
Proportion Marijuana	-0.426	0.021	***	-0.278	0.028	***	-0.240	0.039	***
Proportion Meth	0.120	0.025	***	0.185	0.030	***	0.087	0.042	*
Proportion Other drug	-0.681	0.082	***	-0.543	0.065	***	-0.508	0.064	***
No. filings per judgeship				0.000	0.000	**	0.000	0.000	**
Median time to felony disposition				0.004	0.001	**	0.005	0.002	**
% HS Grad				0.003	0.001	**	0.006	0.002	**
% Poverty				-0.008	0.001	***	-0.004	0.002	**
Violent crime rate				0.000	0.000	***	0.000	0.000	*
Treatment admissions rate				-0.004	0.001	***	-0.003	0.001	*
% Black				0.003	0.001	***	0.003	0.001	**
Density				0.000	0.000	***	0.000	0.000	***
Midwest				0.025	0.006	***	0.028	0.014	*
Northeast				-0.093	0.012	***	-0.112	0.015	***
West				-0.003	0.006		0.012	0.014	
Constant	0.709	0.021	***	0.398	0.117	***	0.117	0.196	
Chi-square	964.766			71958.9			26696.68		
R2	0.258			0.399			0.344		

***p<0.001; **p<0.01; *p<0.05

C. Figures

Figure 1: Sentence lengths by drug type over time

Median sentence by drug
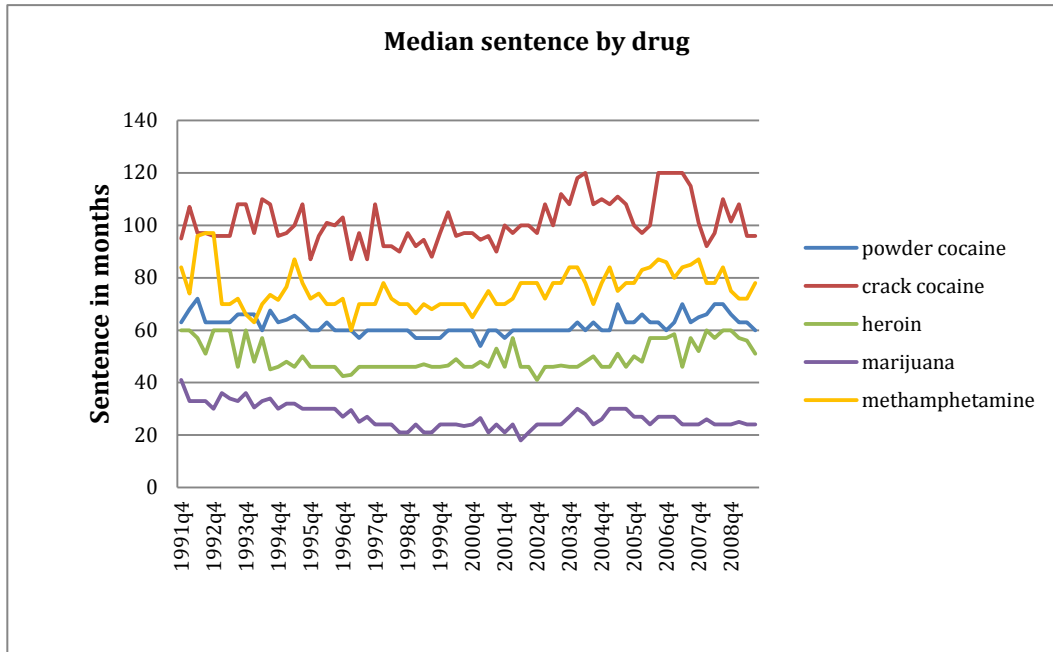
Figure 2: Percent mandatory minimums for drug trafficking cases

Percent mandatory minimums for drug trafficking cases

Legend:
- powder cocaine
- crack cocaine
- heroin
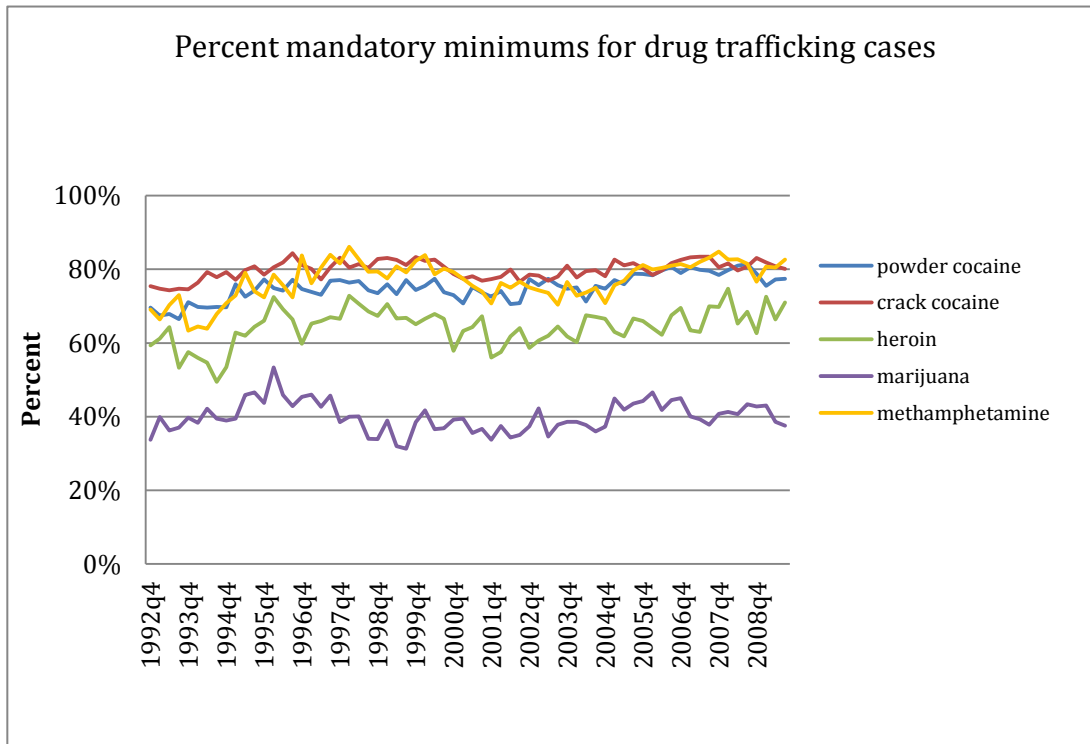- marijuana
- methamphetamine

Figure 3: Mean drug trafficking sentence over time

Mean sentence over time
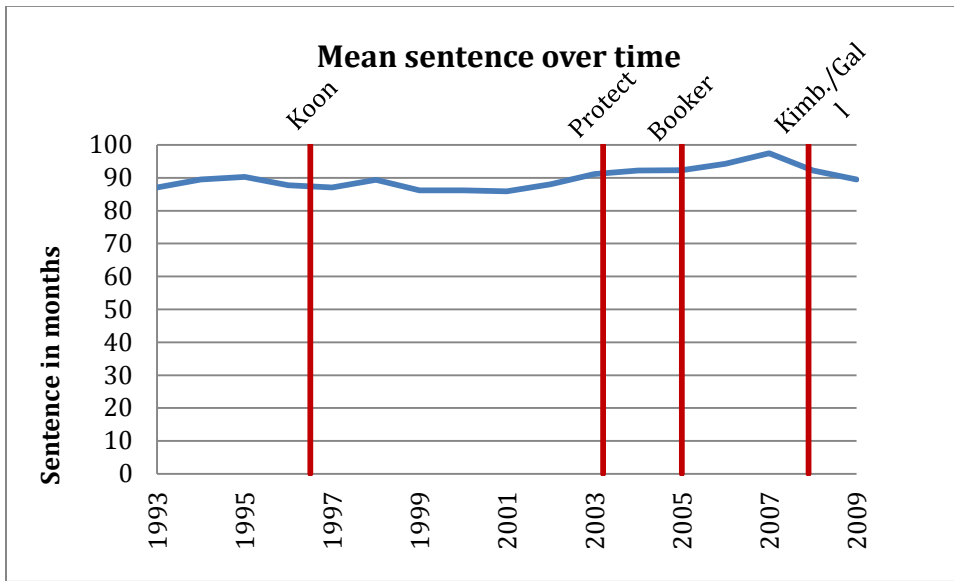
Figure 4: Mean sentence difference for drug trafficking cases over time

Mean sentence difference over time
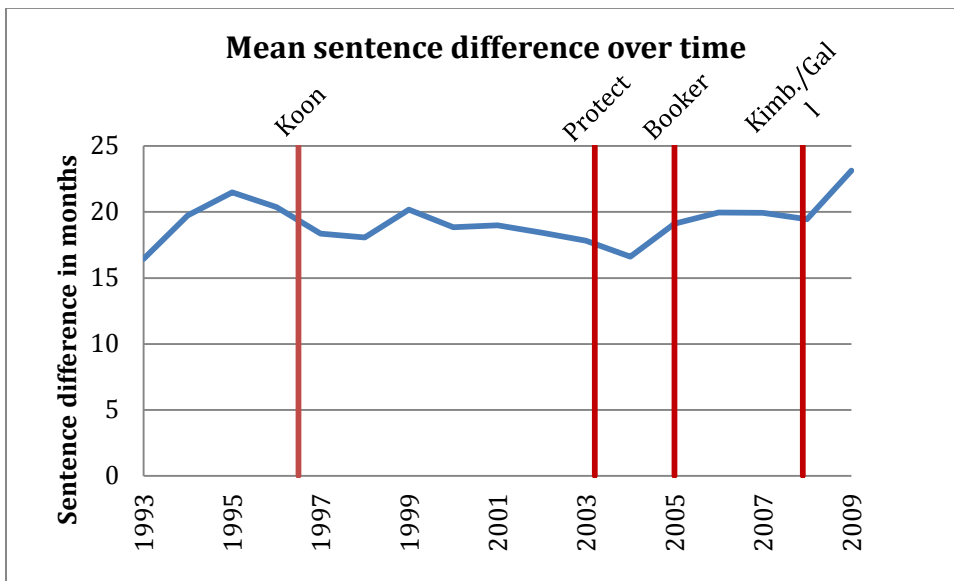
Figure 5: Proportion of sentenced drug trafficking cases subject to a mandatory minimum sentence over time

Percent mandatory minimum in districts over time
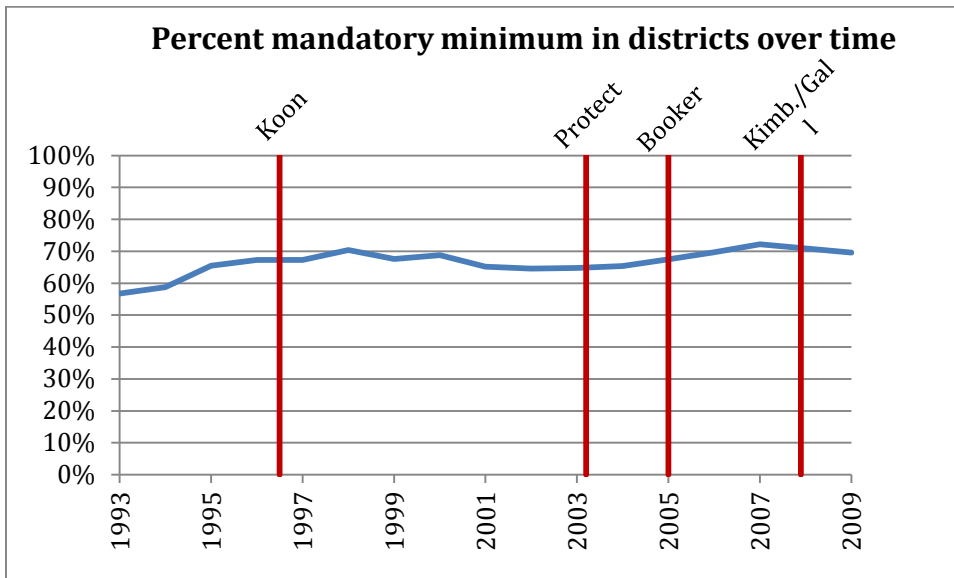
D. Illustrations

Illustration 1: Percentage of Criminal Caseload comprised of Drug Trafficking Cases by District: District Level Maps

Percent drug trafficking cases, 1992-2009

% Trafficking Cases
- less than 25%
- 26% - 35%
- 36% - 45%
- greater than 45%
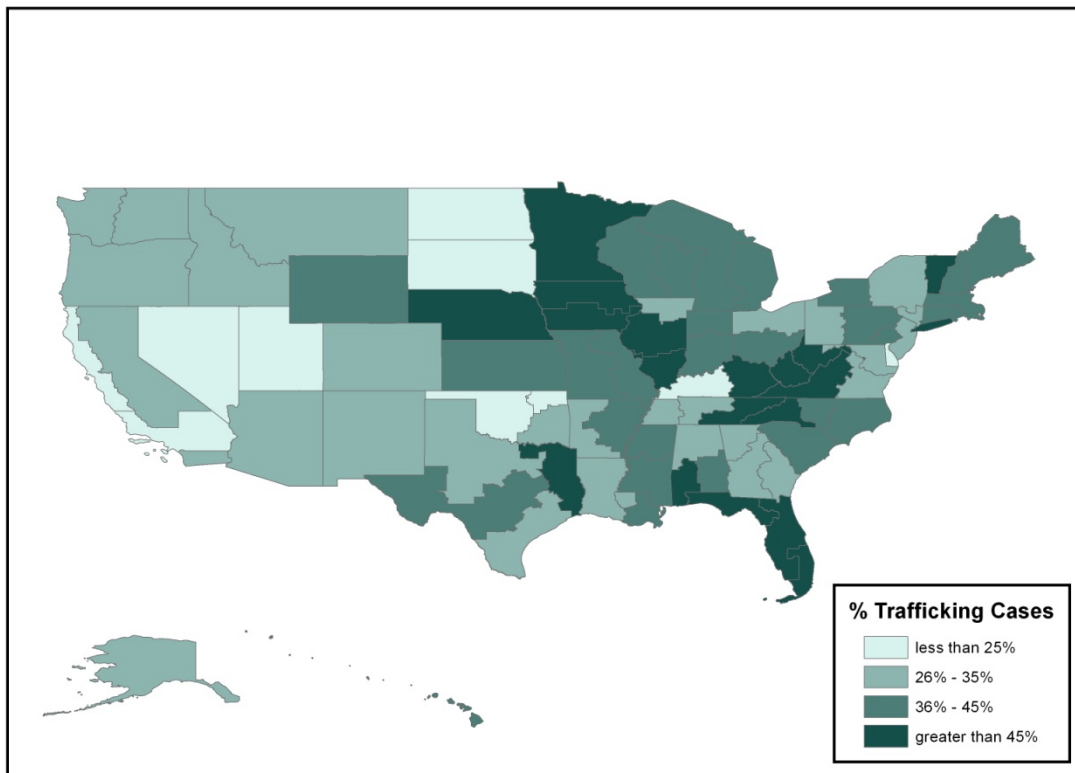
Illustration 2: Relative Distribution of Drug Caseloads, by Drug Types: District Level

Maps

Percent crack cocaine trafficking cases, 1992-2009
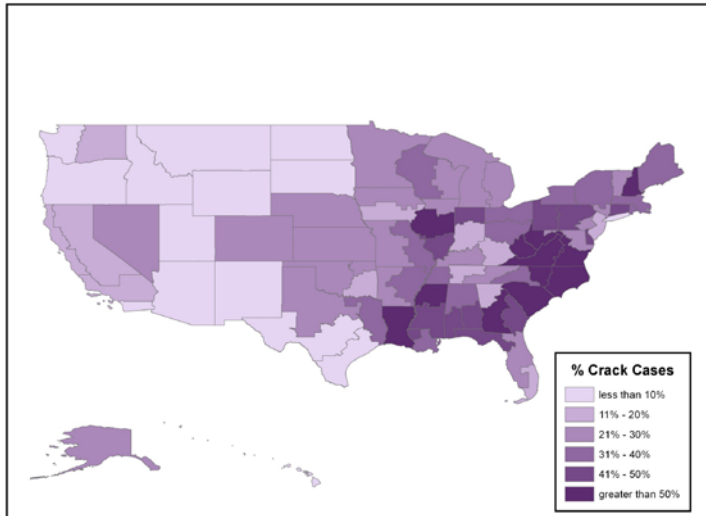
Percent powder cocaine trafficking cases, 1992-2009
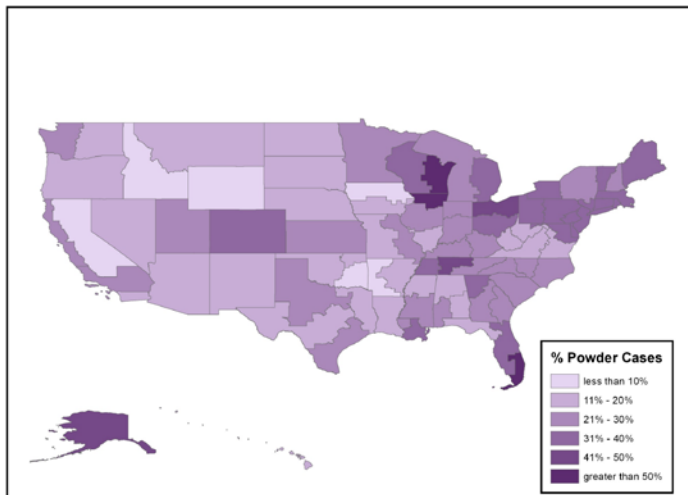
Percent heroin trafficking cases, 1992-2009

% Heroin Cases

less than 10%
11% - 20%
21% - 30%
31% - 40%
41% - 50%
greater than 50%
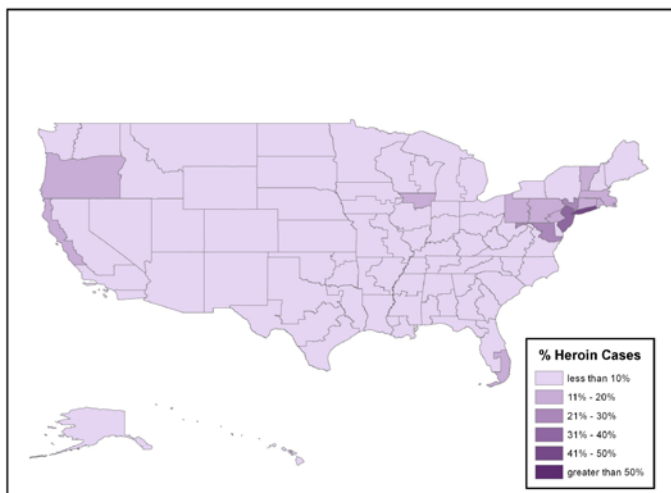
Percent marijuana trafficking cases, 1992-2009

% Marijuana Cases

less than 10%
11% - 20%
21% - 30%
31% - 40%
41% - 50%
greater than 50%
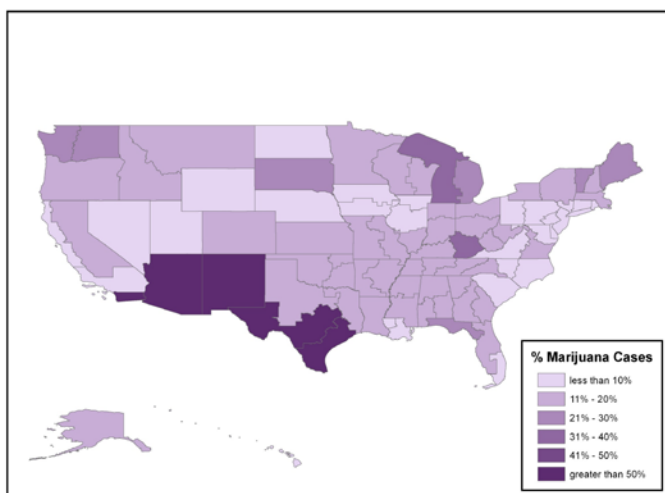
Percent methamphetamine trafficking cases, 1992-2009
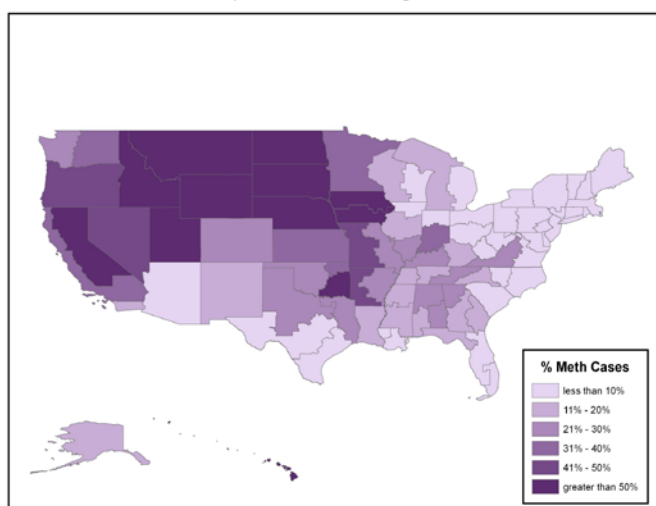
Conclusions:

A. Discussion of findings

Our findings suggest that districts vary considerably from each other in sentencing over time, so the federal system should not be treated as a single, unified system that changes, lockstep, in response to policy mandates. Rather, as Johnson, Ulmer and Kramer (2008) have suggested, district courts reflect local community concerns, district level prosecutors' priorities, and local courtroom workgroup norms, in both caseload qualities and outcomes. As illustrated in the drug trafficking caseload map (Illustration 1), the between-district variability begins with the composition of the drug caseload, where some districts handle relatively few trafficking cases, and others have criminal caseloads in which the majority involve drug trafficking. Within that variability is another layer of variation as to the specific types of illicit substances that are prioritized, as illustrated in Illustration 2.

Second, the data reflect relative stability of outcomes within districts over time, rather than radical change, including in response to the Supreme Court's mandates in *Booker* and *Kimbrough/Gall*. Moreover, drug sentencing is complicated by high, and highly variable rates of mandatory minimum usage by districts, which again is partly driven by specific drug caseload compositions at the district level. As illustrated by Figure 2, powder cocaine, crack cocaine, and methamphetamine all are sentenced under mandatory minimums in about 75-80% of trafficking cases, whereas heroin has slightly lower rates, and marijuana's rate is much lower. Consequently, while across all districts, approximately two-thirds of all drug trafficking cases are sentenced under mandatory minimums, this rate ranges from a low of about 1/3 of the drug trafficking cases in the Southern West Virginia district to 88.4% in the Southern Iowa district. Crack cocaine in particular is heavily enforced in the South, whereas marijuana prosecutions are concentrated along the southwest border states, and methamphetamine is concentrated in the mountain west and parts of the Midwest.

Third, in concurrence with Johnson, et al. (2008), the findings of this study also highlight the importance of considering how sentencing is shaped by a number of meso- and macro-level factors that mediate how policy is put into practice. Thus, sentencing trends transcend policy change, and it appears that while there is some stasis over time in

regard to sentence length (with the possible exception of the PROTECT Act period, which had longer sentences than all of the other periods), the mechanisms for getting to those outcomes adapt to policy mandates. Notably, because prosecutors' tools for decreasing final sentence outcome are subsumed within the Guidelines minimums (especially the granting of substantial assistance departures), there is less observed variation in our sentence difference measure pre-*Booker* than post-*Booker,* but overall outcomes remain fairly stable (see Figures 3-5 for illustration). Moreover, in those districts with high volumes of cases in general, and high proportion of drug trafficking cases, we see a strong tendency toward sentencing norms that persist across time periods, suggesting that the local legal culture is an especially important element to case outcomes above and beyond national level policy mandates in districts that handle large, and drug trafficking-heavy caseloads.

B. Policy Implications

Our findings also call into question the U.S. Sentencing Commission's interpretation of post-*Booker* sentence outcomes, and suggest the importance of using appropriate variable specification, multi-level predictors and controls, and the fullest range of time periods available. We found little evidence that in drug trafficking cases, judges' recently-restored freedom to sentence outside of the Guidelines accounts for increases in observed variations in sentencing at the individual or district levels. In part, this is confirmed by our main effect findings that a higher proportion of mandatory minimum cases predicts more variance from the guideline minimum, and that Guidelines cases are no more likely to have sentence deviations post-*Booker* than are mandatory minimums. Since judges have, throughout the entire period, been constrained from deviating from mandatory minimums, this also suggests prosecutorial behavior is a very important component to such variances.

For policy-makers, this would indicate that reinstating constraints on the judicial sentencing process will not achieve desired uniformity. Thus, enacting additional statutory minimum sentences is not an appropriate policy response in that regard, nor would be other versions that aim to constrain judges. Moreover, the observed regional effects transcend jurisdictional lines, suggesting an important influence of broader

65

cultural norms and historical contingencies in how federal courts operate on the ground, which is not easily constrained by national-level policy mandates.

Finally, from a policy standpoint, the existence of a dual system—a Guidelines system that is designed to take into account a large number of case and offender factors so the final calculation provides for a tailored sentence that still aims for some uniformity, along with a mandatory minimums system that by design ignores many of those key sentencing factors carefully crafted by the U.S. Sentencing Commission and is quite blunt and imprecise---appears to allow for an unintended, but nonetheless exploited, layer of prosecutorial discretion in the federal administration of justice. Thus, our findings, among others, at least indirectly suggest that disparities in the use of mandatory minimums and their corollary mechanisms (such as substantial assistance departures) may be more problematic than changes to the mandatory status of the Guidelines.

Consequently, if federal law-makers are interested in reducing unwarranted sentencing disparities, their best first step might be to consider ways of reining in prosecutorial discretion at the multiple decision-making stages in which they hold considerable or sole power: case selection, initial and final charges, and the power to approve substantial assistance departures, or not. On this final point, it is the very nature of the prosecutor's most wielded departure weapon (substantial assistance) that accounts for some of the most troubling disparities since it is governed not by the relative culpability and deservedness of the sentenced offender, but rather on how much information the sentenced offender can "give" to the prosecutor in exchange for the reduction. Thus, it is not surprising that the most serious drug offenders in our analysis were the ones who benefited the most.

More broadly, as has been suggested before (Engen, 2011; Lynch, 2011a & 2011b; Ulmer, 2012) we suggest that relying only on quantitative analyses of outcome data, such as that collected by the U.S. Sentencing Commission, in order to understand how legal policy impacts the day-to-day functioning of criminal justice systems will provide only a partial "empirical picture" of the answers to key policy questions. Thus, we suggest that studies like this one need to be supplemented by a robust body of empirical research that uses other data sources, takes different methodological approaches, and asks a wider range of research questions to fully uncover the linkages

between formal policy change and its on-the-ground manifestations. In the federal context, this will require increased cooperation by those who work in the system as a way to help achieve the Commission's goal of having an empirically informed system of sentencing (US Sentencing Commission, 2009).

References:

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

Anderson, J., Kling, J., and Stith, K. (1999). Measuring inter-judge disparity: Before and after the federal sentencing guidelines. *Journal of Law and Economics,* 42, 271-307.

Beck, N. (2001). Time-series—cross-section data. *Statistica Neerlandica,* 55(2), 111-133.

Beck, N. and Katz, J.N. (1995). What to do (and not to do) with time-series cross-section data. *The American Political Science Review,* 89(3), 634-647.

Bollen, K. and Barb, K. (1981). Pearson's R and Coarsely Categorized Measures. *American Sociological Review*, 46(2), 232-239.

Doerner, J.K. and 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). Implications of determinate and presumptive sentencing: Making research relevant. *Criminology & Public Policy,* 8, 23-335.

Engen, R. (2011). Racial disparity in the wake of *Booker/Fanfan*: Making sense of "messy" results and other challenges for sentencing research. *Criminology and Public Policy*, 10, 1139-1149.

Everett, R. and Wojtkiewicz, R. (2002). Difference, disparity, and race/ethnic bias in federal sentencing. *Journal of Quantitative Criminology*, 18, 189–211.

Farrell, A. and Ward, G. (2011). Examining district variation in sentencing in the post-

Booker period. *Federal Sentencing Reporter,* 23, 318-325.

Feldmeyer, B., and Ulmer, J. (2011). Racial/ethnic threat and federal sentencing. *Journal of Research in Crime and Delinquency*, 48, 238-270.

Gall v. United States 552 U.S. 38 (2007).

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

Hofer, P.J. (2007). *United States v. Booker* as a natural experiment: Using empirical research to inform the federal sentencing policy debate. *Criminology & Public Policy,* 6, 433–60.

Hofer, P. (2011). Has *Booker* restored balance? A look at data on plea bargaining and sentencing. *Federal Sentencing Reporter,* 23, 326-332.

Johnson, B., Ulmer, J., and Kramer, J. (2008). The social context of guidelines circumvention: The case of federal district courts. *Criminology*, 46, 737-783.

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

Kimbrough v. United States 552 U.S. 85 (2007).

Lynch, M. (2011a). Mass incarceration, legal change and locale: Understanding and remediating American penal overindulgence. *Criminology and Public Policy,* 10*,* 671-698.

Lynch, M. (2011b). Expanding the empirical picture of federal sentencing: An invitation. *Federal Sentencing Reporter,* 23, 313-317.

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.

Nagel, I. and Schulhofer, S. (1992). A tale of three cities: An empirical study of charging and bargaining practices under the federal sentencing guidelines. *Southern California Law Review,* 66, 501-566.

Piehl, A.M. and Bushway, S. (2007). Measuring and explaining charge bargaining. *Journal of Quantitative Criminology,* 23, 105-125.

Rabe-Hesketh, S. and Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.

Rhemtulla, M., Brosseau-Liard, P. & Savalei, V. (in press). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods in non-ideal conditions. *Psychological Methods.* Retrieved from http://www.crmda.ku.edu/people/mijke/files/HowManyCategories.pdf

Schulhofer, S. and Nagel, I. (1989). Negotiated pleas under the federal sentencing guidelines: The first fifteen months. *American Criminal Law Review*, 27, 231-288.

Schulhofer, S. and Nagel, I. (1996). Plea negotiations under the federal sentencing guidelines: Guideline circumvention and its dynamics in the post-*Mistretta* period. *Northwestern University Law Review*, 91, 1284-1316.

Scott, R. (2009). The effects of *Booker* on inter-judge sentencing disparity. *Federal Sentencing Reporter,* 22, 102-108.

Shermer, L. and Johnson, B. (2010). Criminal prosecutions: Examining prosecutorial discretion and charging decisions in U.S. Federal district courts. *Justice Quarterly,* 27, 394-430.

Spohn, C. and Fornango, C. (2009). US Attorneys and substantial assistance departures: Testing for interprosecutor disparity. *Criminology,* 47, 813-846.

Steffensmeier, D. and DeMuth, S. (2000). Ethnicity and sentencing outcomes in U.S. federal courts: Who is punished more harshly? *American Sociological Review 65,*705–29.

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

Stith, K. and Cabranes, J. (1998). *Fear of judging: Sentencing guidelines in the federal courts.* Chicago: University of Chicago Press.

Tiede, L.B. (2009). The impact of the Federal Sentencing Guidelines and reform: A comparative analysis. *Justice System Journal*, 30, 34-49.

Ulmer, J. (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. (2012). Recent developments and new directions in sentencing research. *Justice Quarterly*, 29, 1-40.

Ulmer, J., Eisenstein, J. and Johnson, B. (2010). Trial penalties in federal sentencing: Extra-Guidelines factors and district variation. *Justice Quarterly,* 27(4), 560-592.

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

Ulmer, J. and Light, M. (2011). Beyond disparity: Changes in federal sentencing after *Booker* and *Gall*? *Federal Sentencing Reporter,* 23, 333-341.

Ulmer, J., Light, M. and Kramer, J. (2011a). Racial disparity in the wake of the

Booker/Fanfan decision. *Criminology & Public Policy*, 10(4), 1077-1118.

Ulmer, J., Light, M. and Kramer, J. (2011b). The "liberation" of federal judges' discretion in the wake of the Booker/ Fanfan decision: Is there increased disparity and divergence between courts? *Justice Quarterly*, 28, 799-837.

United States v. Booker 543 U.S. 220 (2005).

U.S. Sentencing Commission. (2006). *Final Report on the Impact of United States v. Booker on Federal Sentencing*. Washington, DC: United States Sentencing Commission.

United States Sentencing Commission (2009). *An Overview of the United States Sentencing Commission.* Washington, DC: United States Sentencing Commission.

U.S. Sentencing Commission. (2010). *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis*. Washington, DC: United States Sentencing Commission.

Ward, G., Farrell, A, and Rousseau, D. (2009). Does racial balance in workforce representation yield equal justice? Race relations of sentencing in federal court organizations. *Law and Society Review,* 43, 757-806.

Wilmot, K. and Spohn, C. (2004). Prosecutorial discretion and real-offense sentencing: An analysis of relevant conduct under the federal sentencing guidelines. *Criminal Justice Policy Review*,15, 324-343.

Worrall, J.L. and Pratt, T.C. (2004). Estimation issues associated with time-series-cross-section analysis in criminology. *Western Criminology Review,* 5, 35-49.

Wu, J., and Spohn, C. (2010). Interdistrict disparity in sentencing in three U.S. district courts. *Crime & Delinquency,* 56, 290-322.

VI. Dissemination of research findings:

The findings of this research have been presented at two scholarly conferences. We presented a talk entitled, "The Changing Landscape of 'Crack'in American Federal Law" at an international, competitive acceptance conference, Beyond the Buzzword: Problematising 'Drugs.' This conference was jointly put on by Monash University in Australia and the academic journal, *Contemporary Drug Problems,* and took place at Monash's conference center in Prato, Italy in October, 2011. In our talk, we presented analyses of longitudinal sentencing data in crack and powder cocaine trafficking cases in the context of policy reforms around crack sentencing.

In November, 2011, we presented preliminary analyses of the main research questions posed in this project at the American Society of Criminology meeting in Washington DC. The panel was organized by Linda Truitt (NIJ), and included papers by all three of the Data Resources program grantees. Our paper was entitled, "Legal change and sentencing norms in federal court: The impact of the *Booker, Gall,* and *Kimbrough* decisions across time & place."

We are currently in the process of writing up two manuscripts that report on the findings from this funded project, and which we plan to submit to peer reviewed journals for publication consideration. The first will report on the main findings (as presented in this technical report) and we hope to submit that paper to *Law and Society Review*. The second paper examines changes to legal policy on crack cocaine sentencing, and as part of that manuscript, we will present some findings form our analyses specific to crack sentencing over time and across districts.