

# TRAC Report: Examining Current Federal Sentencing Practices: A National Study of Differences Among Judges



## SUSAN B. LONG

Co-Director,  
Transactional  
Records Access  
Clearinghouse,  
Syracuse University  
Associate Professor  
of Managerial  
Statistics, Martin J.  
Whitman School of  
Management,  
Syracuse University

## DAVID BURNHAM

Co-Director,  
Transactional  
Records Access  
Clearinghouse,  
Syracuse University  
Associate Research  
Professor, S.I.  
Newhouse School  
of Public  
Communications,  
Syracuse University

An analysis of more than 370,000 criminal cases completed in the nation's federal courts during the last five years has documented extensive and hard-to-explain differences in the sentencing practices by the judges working in many federal districts.

This first-of-its-kind, judge-by-judge review by the Transactional Records Access Clearinghouse (TRAC) at Syracuse University<sup>1</sup> of federal sentences imposed for drugs, white-collar, and other kinds of crimes from FY2007 through FY2011 indicates that the typical sentence handed down by a federal district court judge can be very different than the typical sentences handed down for similar cases by other judges within the same courthouse. Significant unexplained differences in sentences were found to exist within offices that were located in at least three out of four of the nation's federal district courts. This finding raises questions about the extent to which a federal sentence is influenced by the particular judge who was assigned to decide it rather than only the specific facts and circumstances of that case.

**Caution.** A key requirement for achieving justice is that the judges in a court system have sufficient discretion to consider the totality of circumstances in deciding that a sentence in a specific case is *just*. No set of rules, including the Federal Sentencing Guidelines, can substitute for this necessary flexibility.

But a fair court system also requires *equal justice* under the law. This means that the average or median sentences of the judges will not be widely different for similar kinds of cases. So the goal of systematically examining sentences is not to develop a lockstep sentencing system. Rather, the goal is to provide both the courts and the public with detailed information so that they can examine whether justice is being achieved.

This article is divided into four sections. We first discuss the study design, including our research objectives and the data and methods used. Next we present our findings on the existence of statistically significant differences in the median sentences handed down by judges in the same courthouse that we found in three out of four of the nation's federal district courts. We then examine the magnitude of these differences in five illustrative district courts. Finally, we provide details on the magnitude of these differences in all federal districts.

## I. The Study Design

### A. Research Objectives

Over the past four decades a long line of studies have examined whether the principle of equal justice under the law was being served in the federal criminal courts. A key purpose of the Federal Sentencing Guidelines, adopted after the passage of the Sentencing Reform Act of 1984, was to reduce what then was perceived as unwarranted sentencing disparity among federal judges. Until a series of Supreme Court decisions between 2005 and 2007 rendered them effectively advisory, the Federal Sentencing Guidelines had defined the parameters of sentencing requirements and in doing so had reduced the sentencing discretion of federal judges.

Despite a great deal of interest in the question of whether equal justice goals were being met by federal sentencing practices, developing a consensus about this challenge has been difficult. Did unwarranted sentencing disparities exist, and if they did, how large were they? Did the advent of Sentencing Guidelines actually materially decrease unwarranted disparity, or simply alter its nature? Now that the Guidelines are advisory, what if anything has changed? As researchers have long noted, both methodological questions and data deficiencies have meant that answering these questions with any certitude has been surprisingly difficult.

The resulting extensive sentencing literature has been well summarized elsewhere and will not be repeated here. A 2010 article by Ryan W. Scott examining sentencing disparities within one district court, and a 1999 article by Paul J. Hofer, Kevin R. Blackwell, and R. Barry Ruback when the two lead authors were at the U.S. Sentencing Commission, are only two of the many articles that provide useful reviews of the empirical literature in this area.<sup>2</sup>

In many respects, the purpose of TRAC's current research is both *less* and *more* ambitious than past studies. A key focus of many of the past articles has been an attempt to assess whether the alleged unwarranted sentencing disparities are increasing or decreasing. In this case, the TRAC study reported here is decidedly less ambitious because it focuses on trying to answer the question of whether sentencing disparities currently exist. And, where they are found, it attempts to describe and document the nature and extent of these differences.

*Federal Sentencing Reporter*, Vol. 25, No. 1, pp. 6–17, ISSN 1053-9867, electronic ISSN 1533-8363.  
© 2012 Vera Institute of Justice. All rights reserved. Please direct requests for permission to photocopy  
or reproduce article content through the University of California Press's Rights and Permissions website,  
<http://www.ucpressjournals.com/reprintInfo.asp>. DOI: 10.1525/fsr.2012.25.1.6.

Obviously, before it is possible to assess whether there are *changes* in sentencing practices, it first must be determined what the sentencing practices actually are today. Since this foundation is now largely lacking, it is our belief that the very first priority must be to describe recent sentencing patterns. Only when a wider consensus emerges about current sentencing practices, and whether unwarranted sentencing disparities exist, can we hope to determine whether they might be increasing or decreasing since the Federal Sentencing Guidelines became advisory.

We are also convinced that this single goal—documenting current sentencing practices—is far too challenging and difficult to be accomplished by any single study including our own. Concluding that finding judge-to-judge *differences* in typical sentences is synonymous with finding *unwarranted disparity* in sentencing practices requires a very solid basis for this leap in inference.<sup>3</sup> If past research has established anything, it has shown that sentencing practices of a judge cannot be meaningfully reduced to a single generality or even a few simple numbers. In the past, attempts to deal with the underlying complexity in this fashion have not led to consensus in part because the very numbers derived are so dependent upon the particular methodologies used and the assumptions made. Not enough time has been spent in examining the actual details of judge-by-judge sentencing patterns.

It is for this reason that a second goal of TRAC's long-term study has been the development of an accessible database continually updated with current sentencing decisions. We want to make this information available to a wide range of individuals—including scholars, lawyers, judges, court administrators, public interest groups, Congressional staffers and others—to encourage more people to undertake their own independent examinations.<sup>4</sup> It is in this way that the effort involved in our current work is decidedly more ambitious than most past studies.

## B. Data and Methodology

TRAC's effort to identify possible data sources that either alone or in combination might allow this type of research on the sentencing practices of federal judges, as well as permit related investigations of the role of federal prosecutors, began over two decades ago. Finding that no such information was available, TRAC began the painstaking task of compiling information from a range of data sources, including records from the federal courts, from federal

prosecutors, from various other divisions within the Justice Department, and from the Office of Personnel Management (OPM). Our goal was to create a comprehensive national database on judge-by-judge *and* prosecutor-by-prosecutor decisions.<sup>5</sup> This article is based upon these data and covers the sentencing records of over eight hundred federal district court judges and their decisions during the last five years (FY2007 through FY2011).

This study, as others have in the past, takes advantage of the broad institutional promise that criminal cases are assigned on a random basis to judges within the same courthouse.<sup>6</sup> When random assignment occurs, we know by the *law of large numbers* that—assuming a reasonably large number of cases—the composition of the cases each judge hears will be roughly comparable. Just as this methodology based upon random assignment to different groups is the gold standard for determining whether *treatments*, such as the administration of alternative drugs or adoption of new medical procedures, have any real impact on patient outcomes, in a similar fashion this methodology provides the most reliable method for isolating the influence of a particular judge's sentencing practices from the impact of all of the multitude of other factors impacting sentence outcome, including the nature and seriousness of the crime and the defendant's personal history.<sup>7</sup>

To ensure a sufficient number of cases, we limited our analysis to district court judges who sentenced at least 50 defendants. During fiscal years 2007 through 2011, 882 federal district court judges met this standard.<sup>8</sup> Six of these judges were commissioned to serve in more than one federal district. For these judges, we examined their caseload separately within each district. This gave us 888 judge comparisons. Each of these 888 judge comparisons involved on average over 400 defendants. See Table 1.

To provide additional controls, we also looked at two more restrictive subsets of these judges. In Subset 1 we eliminated any judge not on the bench for the entire 60-month period we examined. We refer to this subset as “continuous service” judges. This helps make sure that the composition of judge caseloads wasn't affected by timing issues of when a judge served, should the nature of the workload at a courthouse have changed over time.

For Subset 2, in addition to focusing on the judges who had served the full 60-month period, we eliminated any judge who had retired at any point prior to the end of fiscal year 2011 and became a senior judge. We refer to this

**Table 1**  
**Federal District Court Judges and Sentences Included in the Study, FY2007–FY2011**

Judge Category*	Qualifying Judges	Defendants Sentenced	Average # Sentenced
Total	888	371,826	419
Continuous Service (Subset 1)	751	342,130	456
Active Judges (Subset 2)	469	243,736	520

\* Total includes all federal district court judges who sentenced 50 or more individuals in a district during FY2007–FY2011. Subset 1 excludes anyone who did not serve during the entire 60 months of the study period. Subset 2 requires service for the entire 60 months and excludes those who retired at any point before the end of the study period.

**Table 2**  
**Federal Judicial Districts and Courthouses**  
**Included in the Study**

Judge Category*	Number Covered**	
	Courthouses	Judicial Districts
Total	154	87
Continuous Service (Subset 1)	137	85
Active Judges (Subset 2)	93	67

\* See Table 1 footnote for definition of qualifying judges in these three nested categories.

\*\* Required that there be two or more qualifying judges to make inter-judge sentencing comparisons.

second subset as “active judges.” Although many retired judges hear large numbers of cases, to the extent senior judges exercise some control over the types of cases they hear, this approach eliminates that potentially confounding factor.

Details of the number of judges in each of these two subsets and the number of defendants they sentenced are also summarized in Table 1. For example, the 469 judges in the “active judges” subset made up just over half (53 percent) of the original total of 888. However, they handled two thirds (66 percent) of the 371,826 cases that were processed during this five-year period, sentencing on average 520 individuals per judge.

We next examined the courthouse where each of these judges was located. For example, for district court judges from the Northern District of Illinois, we distinguished between the Chicago and Rockford offices. We did this in each district court. Only courthouses that had at least two judges who met our other qualifications were included in our analyses.<sup>9</sup> Our entire sample of judges gave 154 offices in 87 federal judicial districts where sentencing patterns could be compared. As we used more restrictive criteria for identifying judges, the number of offices in which comparisons were possible fell. As shown in Table 2, given the important role senior judges play, the principal reduction occurred when we restricted the analysis to “active judges.” But even here, a total of 93 offices in 67 separate federal districts could be included in our comparisons.<sup>10</sup>

### C. Using Average vs. Median Sentences

Sentence length is usually compared in terms of average or median sentences. For the median value, the middle sentence in length is selected: half are shorter and half are longer. Most past studies utilizing random assignment methods have been based on the general linear model and used techniques such as the analysis of variance (ANOVA). For these, calculations involve means (averages) and variances (the average of squared distances of sentence lengths from the average). Since the median and average values will be approximately the same in data that is symmetrically distributed, there is usually no need to look at both types of measures with many kinds of data. Looking only at average values is sufficient.

However, sentencing data generally is *not* symmetrically distributed. Rather, it is what statisticians refer to as “skewed,” where the right-hand side of the distribution stretches out to encompass a range of longer sentences. Accordingly, the issue of whether to use the average sentence or the median sentence is an important decision in sentencing research.<sup>11</sup>

In addition, *outliers* are frequently present—that is, one or more extremely long sentences as compared with the rest of the sentences given. This further compounds the difficulties. Skew and the presence of outliers each result in the *average* being much higher than the *median* sentence. Not only do skew and outliers affect the value of the mean, they have an even greater impact on estimates of the variation among sentence lengths because these are based upon squared differences. Estimates of the size of sentencing disparity based upon average variance explained and other related measures are thus very sensitive to the presence of skew and outliers. As a result, these measures may give an inflated picture of any actual differences among judges in sentencing patterns. That is, calculations based on means and variances that assume normally distributed data without outliers can result in *biased* (wrong) estimates, as well as in false results from statistical tests.

Before basing our analyses on ANOVA methods, we therefore examined whether the normality assumptions underlying the use of mean values were satisfied by our current sentencing data. Both standard tests for normality, as well as graphical methods examining the fit of the residuals after ANOVA was run on these data, indicated that the use of average sentences as our measure was not appropriate because the underlying assumptions required by the model were not met. Not only were there consistent statistically significant departures from normality, office-by-office, in the residuals, but also the magnitude of these departures resulting from skewness were quite pronounced.

Thus, to guard against reaching inaccurate conclusions and to derive conservative estimates of any differences in sentence length among judges, TRAC based its analysis on median sentences. Kruskal-Wallis tests were used to determine whether differences among judges in their median sentences were statistically significant (that is, they cannot be reasonably attributable to chance variations).

## II. Significant Differences Found Among Judges in the Length of Sentences Given

Three sets of analyses were run comparing median sentences of the judges meeting the criteria outlined earlier, who were based in each of the courthouses noted in Table 2. First, the analysis was run using all judges who had sentenced at least 50 defendants during the time period covered by this study. Second, the analysis was rerun using the subset of judges who had served for the entire period. Finally, the analysis was run once more on only the “active judges,” omitting any judges who had retired at any point before the end of the study period.

**Table 3**  
**Number of Federal Courthouses and Districts with Significant Differences in Median Sentence Length**

Judge Category*	Total Comparisons**	Statistically Significant Difference***	
		Number	Percent
Courthouse Comparisons			
Total	154	95	62%
Continuous Service (Subset 1)	137	87	64%
Active Judges (Subset 2)	93	52	56%
Districts with One or More Statistically Significant Courthouses			
Total	87	68	78%
Continuous Service (Subset 1)	85	65	76%
Active Judges (Subset 2)	67	41	61%

\* See Table 1 footnote for definition of qualifying judges in these three nested categories.

\*\* Required that there be two or more qualifying judges to make inter-judge sentencing comparisons.

\*\*\* Statistically significant at the .05 level or lower for Kruskal-Wallis test on inter-judge differences in median sentence length.

Table 3 summarizes the results from these tests. Around six out of ten courthouses were found to have statistically significant differences in the median sentences handed down by judges on those benches. One or more courthouses with statistically significant sentencing differences were located in approximately three out of every four federal districts. When only the “active judges” subset was examined, the proportion of courthouses with statistically significant differences was somewhat lower, with 56 percent of them showing a statistically significant difference. These courthouses were located in 61 percent of the federal districts in which comparisons could be made.

What do these results mean? Assuming that cases were randomly assigned to judges located in that same courthouse, finding statistically significant differences in median sentences implies that we can reasonably rule out as an explanation of the observed differences any potential variation in the makeup of the crimes, the defendants, or other factors influencing outcome randomized across judges in the assignment process. Having ruled out these alternative explanations, we are left with the strong possibility that different sentencing philosophies or practices among judges may help explain these differences.

However, the fact of finding statistically significant differences does not tell us whether these differences are particularly large or small—sentences differing by a few weeks, a few months, or many years. Statistical significance is a function of both the number of observations (e.g., the defendants each judge sentenced) and the magnitude of differences in sentence lengths. Given that in this study each judge on average sentenced hundreds of defendants, these differences could have been fairly small but still able to rule out *chance variability* in the makeup of cases as the explanation.<sup>12</sup>

Accordingly, in the remaining sections of this paper, we provide details on the character and magnitude of the sentencing differences found.

### III. Judge-to-Judge Variation in Sentencing: Some Illustrative Districts

To help place the magnitude of sentencing differences into a more easily understood context, we next examined how many months the typical or median sentences of judges differed for the two largest classes of federal crimes generally handled by district court judges: drug offenses and white-collar crimes. For these comparisons, in an abundance of caution, we imposed an additional requirement on the judges who were compared to assure that not only had they sentenced a sufficiently large number of defendants in general, but also sentenced a large number (at least 40) individuals convicted of the class of crime we examined.<sup>13</sup>

To be conservative, we again used each judge’s median sentence for these comparisons because this measure is much less variable and little influenced by particularly long sentences that might on occasion arise when by chance a judge is assigned some especially unusual cases to handle.

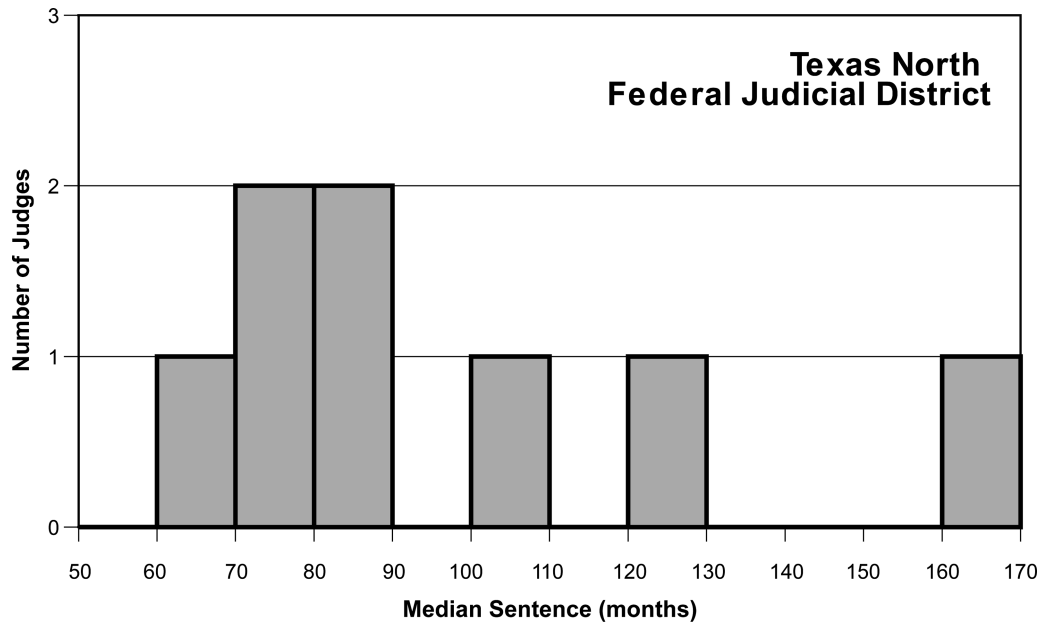
Here are a few examples involving sentencing differences for drug convictions:

**Northern Texas Judges.** In the Northern District of Texas statistically significant overall sentencing differences were found for the judges in both Dallas and Fort Worth. The median or typical sentences for drug cases during the past five years were then calculated for each Northern Texas judge. The comparison, each of whom had handed down at least 40 drug sentences, is shown in Figure 1.

At the low end of this small group of eight federal judges, the median sentence was 60 months. At the other extreme, however, was a judge whose median sentence was 160 months. Nor did the sentences of the remaining six judges cluster together. In fact, their typical sentences were quite varied.

The median drug sentences for the four judges located in Dallas ranged from 60 months for the lowest judge to 121.5 months for the highest judge,

**Figure 1**  
**Texas North Judges: Median Drug Prison Sentences**



or a sentence twice as long. The median sentence for one of the two judges located in Fort Worth was 160 months, or over 50 percent higher (nearly 5 years longer) than the other judge, whose median sentence was 102.5 months. Assuming the drug cases handled by these judges were assigned on an approximately random basis, this variation is hard to explain.

No statistical tests could be performed for the other two offices in the district because each only had a single judge meeting our criteria. The median drug sentence was 87 months for both the judge in Amarillo and the judge in Lubbock.

**Eastern Virginia Judges.** In the Eastern District of Virginia, which includes Alexandria, Norfolk, and Richmond, fifteen judges had each sentenced at least 40 drug offenders. Each of these offices showed statistically significant overall sentencing differences on our tests. As shown in Figure 2, the range in the typical sentence here was again very broad. The median sentence for three of the judges was 120 months, four times that of the median sentence of 30 months from the judge at the other end of the scale.

Part of these differences might be accounted for by differences in the composition of cases assigned judges located in different offices within the district. Alexandria judges had median sentences that ranged from a low of 30 months to a high of 87 months, whereas Norfolk judges ranged from 79.5 months to 120 months. In the case of Richmond, median sentences ranged from 78 months to 111 months. Thus, even within each office, the range in median sentences was still large.

Although large district differences were not uncommon, in some districts there was relatively strong agreement in the sentencing practices of the judges.

**Minnesota Judges.** The Minnesota District Court was one location where fairly close agreement among judges on sentencing occurred. There were no significant differences found in our tests for Minneapolis judges.<sup>14</sup>

As shown in Figure 3, the nine active and senior district court judges there who had handled at least 40 drug cases clustered closely on their median sentences, with a low of 52 and a high of 64 months. The range for Minneapolis judges was between 54 and 64 months, and St. Paul judges ranged between 52 and 60 months.

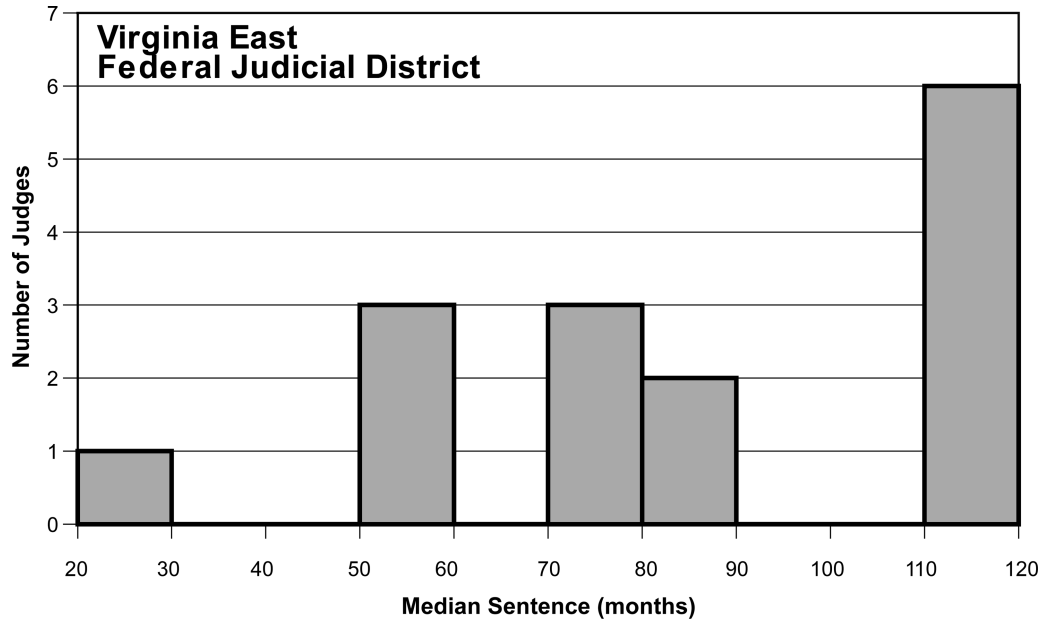
Districts with two or three judges often showed more agreement but, this wasn't always the case.

**District of Columbia Judges.** In the District of Columbia federal court, all judges are located in the same courthouse, and statistically significant sentencing differences were found in our tests. In that courthouse only three judges had handled at least 40 drug cases. The shortest median was only 27 months, the longest was 77 months—almost three times longer—and the third, at 51 months, fell between them. See Figure 4.

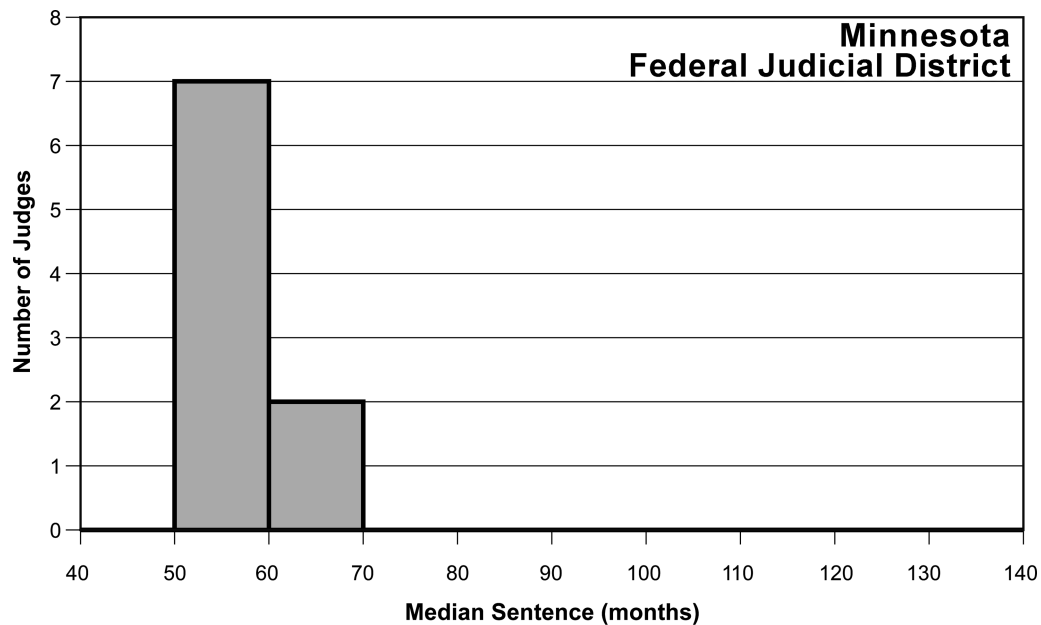
Similarly varied patterns were observed in the judge-by-judge sentences for white-collar crime:

**Northern Illinois Judges.** In the Northern District of Illinois, statistically significant sentencing differences were found for Chicago judges. Here the records showed a total of eight judges, all based in Chicago, who had sentenced 40 or more defendants for white-collar crimes from FY2007 through FY2011. The median sentence of these eight judges ranged from a low of zero—that is, at least half of the defendants before that judge received no prison time—to a high of 39 months. See Figure 5.

**Figure 2**  
**Virginia East Judges: Median Drug Prison Sentences**



**Figure 3**  
**Minnesota Judges: Median Drug Prison Sentences**



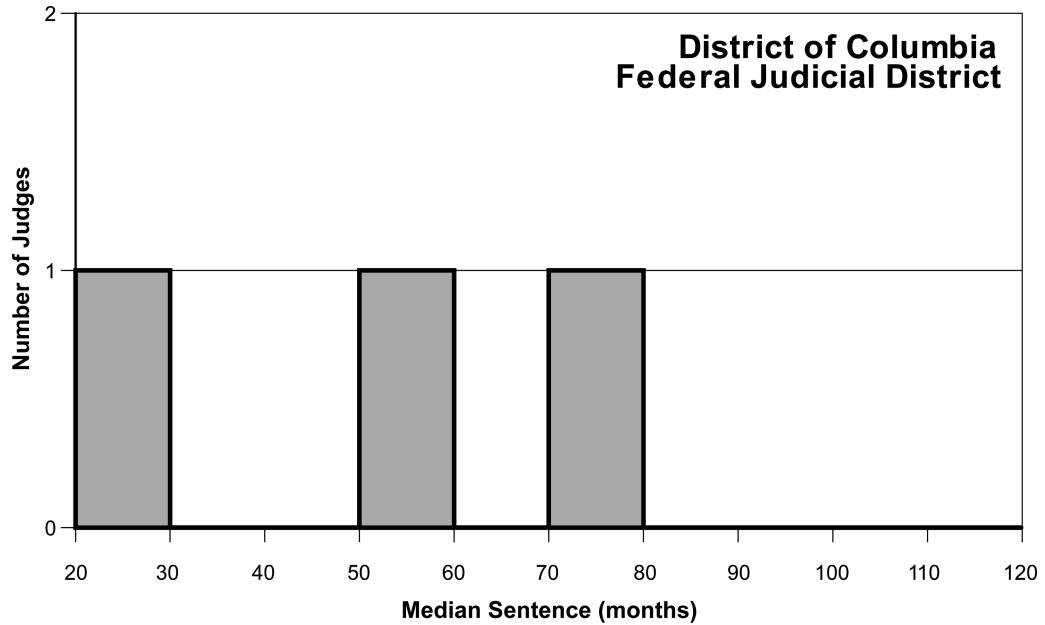
#### **IV. The Magnitude of Sentencing Differences in All Federal Districts**

After examining sentencing differences in a handful of illustrative districts, we expand our focus to look at the sentencing patterns in all federal judicial districts where such comparisons were possible. Two areas were explored. First, what was the magnitude of these judge-to-judge differences in median sentences? Second, to what extent did

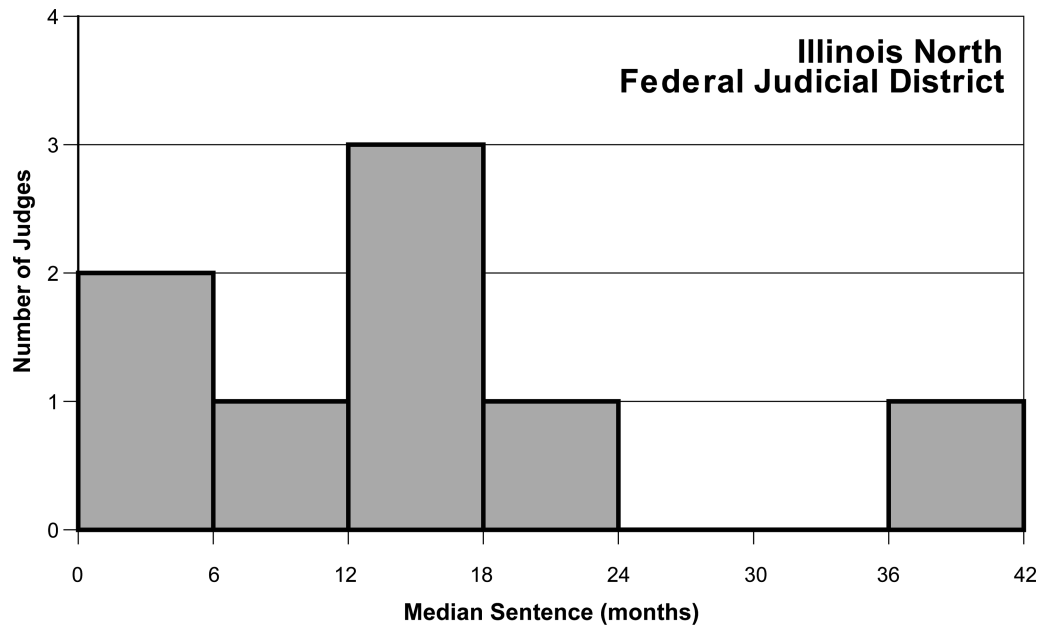
the differences *within* the individual courthouses account for the differences in judge-to-judge median sentences for the *district* where they were located?

There was a great deal of variability among the districts with regard to these questions. As noted earlier, the data show that in four out of ten of the *courthouses*, no statistically significant judge-to-judge differences in median sentences were found. And a quarter of the *federal districts*

**Figure 4**  
**District of Columbia Judges: Median Drug Prison Sentences**



**Figure 5**  
**Illinois North Judges: Median White-Collar Prison Sentences**



had no office in which statistically significant differences were discovered. But even where statistically significant differences existed, the magnitude of the inter-judge differences varied from one district to the next. Quite simply, there are no easy ways to encapsulate the wide diversity we found. The best we can hope to do is to convey some of the broad features.

We start first with a brief description of the patterns in the sentences imposed in drug cases, and then look at

differences in the median sentences for white-collar offenders.

#### A. Drug Sentences: Judge-to-Judge and Courthouse-to-Courthouse Differences

The extent of judge-to-judge differences in the median drug sentences within federal courthouses across the country was highly varied. In one courthouse (Atlanta), the judge-to-judge median sentences differed by ninety months. In nine

other courthouses, median drug sentences of the judges agreed, and there were no differences. Between these two extremes were twenty-nine other courthouses with sentencing differences of three years or more, and fifty-two with differences of at least two years.

The entire distribution of the magnitude of these inter-judge differences *by courthouse* is shown in Table 4. The difference measures were calculated as follows. Within each courthouse, the median drug sentence for each judge was ordered from lowest to highest. The difference was then calculated by subtracting the median drug sentence for the lowest judge *at that courthouse* from the corresponding median sentence for the highest judge located *in that same courthouse*. From the results displayed in Table 4, it seems fair to conclude that when differences in median sentence lengths were statistically significant, they were often sizable differences.

Table 4 also shows the distribution of judge-to-judge differences in the median drug sentences for all judges within *the same district*. These measures were calculated in a similar fashion, but here the median drug sentence for the lowest judge *in the district* was subtracted from the corresponding median sentence for the highest judge located *in that same district*. Because most districts have more than one courthouse, one would expect the variability through the district to be higher than within any individual courthouse. This is because in part different courthouses may use different random assignment pools, and thus the composition of drug cases handled by judges from different offices may accordingly differ. In addition, sometimes differences within a district also may be related to distinctive sentencing philosophies in different courthouses.

Surprisingly, the range of differences documented in the *districts* was only slightly wider than that of individual *courthouses*. The largest district judge-to-judge difference found in the median drug sentence was 100 months, whereas for courthouses it was 90 months. The maximum district difference occurred in the Northern District of Texas. At the other extreme was a small district, Rhode Island, where the median drug sentence for the two judges was identical.

When all courts were compared, the typical *district* judge-to-judge difference was 31.5 months. By contrast, the typical difference in the sentences imposed within the 133 *courthouses* was just 18 months, or 57 percent of that found for the districts. In addition, 38 districts, versus 29 courthouses, had sentencing differences of three years or more. And 56 districts, versus 52 courthouses, had sentencing differences of at least two years. Acknowledging that districts vary in their patterns, it still seems fair to conclude that for many districts the majority of the range observed was within courthouses and thus presumably within the same random assignment pool.

Table 5 displays the ten *courthouses* with the largest differences among judges on their median drug sentence.

**Table 4**  
**Differences in Drug Sentences: Comparing the Distribution of Courthouses versus Districts\***

Judge-to-Judge Differences	Number	
	Within Courthouses	Within Districts
8 years or more	0	1
7 years or more	1	4
6 years or more	2	6
5 years or more	5	14
4 years or more	12	21
3 years or more	29	38
2 years or more	52	56
18 months or more	69	69
12 months or more	96	79
6 months or more	113	84
<i>Total Comparisons</i>	133	88
<i>Differences (months):</i>		
<i>median</i>	18.0	31.5
<i>lowest</i>	0	0
<i>highest</i>	90	100

\* Difference between the median drug sentence of the lowest versus highest judge in the respective district or courthouse.

Table 6, on the other hand, shows the ten *districts* with the largest differences. (Because five districts tied for 10th place, there are actually fourteen districts on the “top-ten” list.)

When comparing the two lists, one notable feature is the considerable overlap in the districts named. In fact, the seven courthouses with the most disparities were all in districts that made the top-ten district list. The data also showed that for ten out of the fourteen districts listed in Table 6, the intra-courthouse differences among judges accounted for over half of the difference in median sentences when all judges within that district were considered.

But there were exceptions. In the districts of Kansas and the Northern District of Florida, for example, no courthouse had a statistically significant difference in median sentences. Yet they still made the top-ten district list because of the considerable difference in the median drug sentences among judges from *different* courthouses. Why these courthouse-to-courthouse differences were so large in these two districts is beyond the scope of the present study because those differences involve judges presumably from different random assignment pools.<sup>15</sup>

As might be expected, courthouses with more judges had a greater likelihood of showing greater inter-judge sentencing differences. But large differences were by no means the exclusive preserve of large courts. One two-judge courthouse made the top-ten list, Ft. Worth, and one out of five of the courthouses in the top twenty involved two-judge comparisons. Half of the courthouses in the top twenty had four or fewer judges. Fairly close agreement was also found in courthouses with a sizable number of judges. San Diego (13 judges), St. Louis (8 judges), Hato Rey in Puerto Rico (8 judges), Minneapolis (7 judges), and Sacramento (7 judges)



**Table 5**  
**Courthouses with the Largest Differences Between Median Drug Sentences\***

Courthouse			Median Sentence (months)			
Ranking	Name	Federal Judicial District	Median Judge	Lowest Judge	Highest Judge	Difference (months)
1	Atlanta	Georgia North	104.8	54.0	144.0	90.0
2	Chicago	Illinois North	72.5	48.0	120.0	72.0
3	Dallas	Texas North	70.5	60.0	121.5	61.5
4	New Orleans	Louisiana East	85.5	60.0	120.0	60.0
5	Philadelphia	Pennsylvania East	50.0	36.0	96.0	60.0
6	Fort Worth	Texas North	131.3	102.5	160.0	57.5
7	Alexandria	Virginia East	60.0	30.0	87.0	57.0
8	Oklahoma City	Oklahoma West	60.0	36.0	91.5	55.5
9	Orlando	Florida Middle	63.0	60.0	112.0	52.0
10	Los Angeles	California Central	50.0	36.0	87.0	51.0

\* Difference between the median drug sentence of the lowest versus highest judge in the respective courthouse.

**Table 6**  
**Federal Judicial Districts with the Largest Differences Between Median Drug Sentences\***

Federal Judicial District		Median Sentence (months)			Judge-to-Judge Difference		Differences statistically significant (based upon all sentences)*
Ranking	Name	Median Judge	Lowest Judge	Highest Judge	Within District	Within Courthouse	
1	Texas North	87.0	60.0	160.0	100.0	61.5	Dallas, Fort Worth
2	Georgia North	101.3	54.0	144.0	90.0	90.0	Atlanta, Atlanta-Gainesville
3	Virginia East	86.0	30.0	120.0	90.0	57.0	Alexandria,** Norfolk, Richmond
4	Florida North	120.0	51.5	136.0	84.5	38.5	**
5	Texas West	48.0	12.0	92.0	80.0	31.0	Austin, El Paso, San Antonio
6	Illinois North	73.0	48.0	126.0	78.0	72.0	Chicago
7	Florida Middle	82.5	51.0	120.0	69.0	52.0	Jacksonville, Orlando, Tampa**
8	Louisiana West	84.0	58.5	120.0	61.5	36.0	Lafayette**
9	Michigan East	46.0	24.0	85.0	61.0	38.5	Detroit
10-tie	Kansas	60.0	48.0	108.0	60.0	22.5	**
10-tie	Louisiana East	85.5	60.0	120.0	60.0	60.0	New Orleans
10-tie	Nebraska	70.0	60.0	120.0	60.0	14.0	Omaha***
10-tie	Pennsylvania East	49.5	36.0	96.0	60.0	60.0	Allentown, Philadelphia
10-tie	South Carolina	93.0	60.0	120.0	60.0	46.0	Charleston, Columbia, Florence, Greenville

\* Difference between the median drug sentence of the lowest versus highest judge in the respective district or courthouse.

\*\* Alexandria is not significant for active judges. For Florida North (Gainesville), probability level .088. Tampa was significant only for all judges, not significant for continuous service and active judge tests. Lafayette was not significant for overall sentences, but did show a significant difference in drug sentences. For Kansas, both Kansas City and Wichita had some probability levels below .10, but above .05.

\*\*\* Differences for median overall sentences was 27 months and highly statistically significant in contrast to only 14 months when comparisons were restricted to only drug sentences.

all had differences of 12 months or less. In St. Louis the difference was only 3 months.

#### B. White-Collar Sentences: Judge-to-Judge and Courthouse-to-Courthouse Differences

The general pattern of results that emerged from our comparison of white-collar sentences was similar to that found for drug sentences. Since fewer white-collar than drug cases were handled by federal judges, fewer courthouses and districts had a sufficient number of cases to permit comparison. Nonetheless it was possible to compare median sentences for white-collar offenders in 72 courthouses located in 65 federal judicial districts.

Again, we see a lot of variability among courthouses—some with considerable difference in the typical sentences of judges dealing with white-collar offenders, whereas for

other judges there was close agreement. In one courthouse, Chicago, the judge-to-judge median sentences differed by 39 months. In four other courthouses, the median white-collar sentences of the judges agreed, and there were no differences. Between these two extremes were 12 courthouses with sentencing differences of 18 months or more and 30 with differences of at least 12 months. The entire distribution of the magnitude of these inter-judge differences *by courthouse* is shown in Table 7.

Although the magnitude of these differences is smaller than those for drug sentences, the difference still amounted to at least an additional year in sentence length in more than four out of ten courthouses compared. In contrast, the typical judge's white-collar sentence in the nation was only 11.5 months. Thus, in relative terms, these

**Table 7**  
**Differences in White-Collar Sentences: Comparing the Distribution of Courthouses versus Districts\***

Judge-to-Judge Differences	Number	
	Within Courthouses	Within Districts
3 years or more	1	1
2 years or more	3	6
18 months or more	12	19
12 months or more	30	38
6 months or more	49	48
Total Comparisons	72	65
Differences (months):		
median	10.5	12
lowest	0	0
highest	39	39

\* Difference between the median white-collar sentence of the lowest versus highest judge in the respective district or courthouse.

differences for judges within *the same courthouse* were also sizable.

Table 7 also shows the distribution of judge-to-judge differences in median white-collar sentences for all judges within *the same district*. Again, this was calculated by taking the median white-collar sentence for the lowest judge in *the district* and subtracting it from the corresponding median sentence for the highest judge located in *that same district*.

Again we see that most of the range within a *district* appears to be accounted for by the differences in median sentence length among judges within *the same courthouse*. Indeed the maximum range in sentence length was identical. Sentence differences for both districts and courthouses varied from 39 months to 0 months. In 19 districts, versus 12 courthouses, had sentencing differences of eighteen months or more. And in 38 districts, versus 30 courthouses, had sentencing differences of at least 12 months.

Table 8 displays the ten *courthouses* with the largest differences among judges on their median white-collar sentence. Table 9, on the other hand, shows the ten *districts* with the largest differences.

When comparing the two lists, we again note considerable overlap in the districts mentioned. Seven out of 10

courthouses were in districts that made the top-ten district list. The data also showed that for nine out of the ten districts listed in Table 9, the intra-courthouse differences among judges accounted for half or more of the difference in median sentences when all judges within that district were considered. And this was true even though each of these districts had multiple courthouses.

## V. Conclusion

This analysis examined more than 370,000 sentencing decisions of over eight hundred federal district court judges in federal courthouses throughout this country over the past five years. The study found statistically significant, unexplained differences in the typical sentences of judges in around 60 percent of these courthouses, while the remaining 40 percent showed no significant differences. It also documented the size of these differences and found many were sizable. Few parts of the country were unaffected. Courthouses with significant differences were located in roughly three out of four federal judicial districts.

As we noted at the outset, documenting sentencing practices is far too challenging and difficult a task to be accomplished by any single study, including our own. However, the results of this study should raise important concerns.

Yet answering the question of whether significant intra-judge differences in sentencing practices exist is not sufficient to establish that such differences are indeed *unwarranted sentencing disparities*. Much more research and a great deal more time is needed for a thorough examination of the actual details of judge-by-judge and prosecutor-by-prosecutor sentencing patterns. Now that TRAC has developed and is making available<sup>6</sup> extensive sentencing data identifying both the judge and prosecutor involved in each case, it is our hope that more scholars, court administrators, and others will take up this challenge. Also required for progress on this front are the informed insights of those who know first-hand the inner workings of our federal trial courts—judges, prosecutors, defense attorneys, and others intimately involved in the work of delivering justice on a day-to-day basis.

**Table 8**  
**Courthouses with the Largest Differences Between Median White-Collar Sentences\***

Courthouse			Median Sentence (months)			
Ranking	Name	Federal Judicial District	Median Judge	Lowest Judge	Highest Judge	Difference (months)
1	Chicago	Illinois North	15.3	0.0	39.0	39.0
2	Baltimore	Maryland	24.0	1.0	27.0	26.0
3	Portland	Oregon	15.0	0.0	24.0	24.0
4	Atlanta	Georgia North	32.0	16.5	40.0	23.5
5	Grand Rapids	Michigan West	24.5	13.0	36.0	23.0
6	Manhattan	New York South	7.5	0.0	22.5	22.5
7	Harrisburg	Pennsylvania Middle	3.0	0.0	21.0	21.0
8	Houston	Texas South	28.3	18.0	38.5	20.5
9	Tampa	Florida Middle	10.0	4.0	24.0	20.0
10	Kansas City	Missouri West	8.0	0.0	19.5	19.5

\* Difference between the median white-collar sentence of the lowest versus highest judge in the respective courthouse.

**Table 9**  
**Federal Judicial Districts with the Largest Differences Between Median White-Collar Sentences\***

Federal Judicial District		Median Sentence (months)			Judge-to-Judge Difference		Differences statistically significant (based upon all sentences)*
Ranking	Name	Median Judge	Lowest Judge	Highest Judge	Within District	Within Courthouse	
1	Illinois North	15.3	0.0	39.0	39.0	39.0	Chicago
2	Maryland	18.0	1.0	28.0	27.0	26.0	Baltimore, Greenbelt
3	Georgia North	34.0	16.5	41.5	25.0	23.5	Atlanta, Atlanta-Gainesville
4	North Carolina West	18.3	0.0	24.0	24.0	11.5	Charlotte
4	Oregon	15.0	0.0	24.0	24.0	24.0	Portland**
4	Texas West	12.0	0.0	24.0	24.0	12.0	Austin, El Paso, San Antonio
7	Michigan West	24.5	13.0	36.0	23.0	23.0	Grand Rapids**
8	New York South	7.0	0.0	22.5	22.5	22.5	Manhattan**
9-tie	Florida Middle	10.0	2.6	24.0	21.5	20.0	Jacksonville, Orlando, Tampa**
9-tie	Texas North	15.0	7.0	28.5	21.5	17.0	Dallas, Fort Worth

\* Difference between the median white-collar sentence of the lowest versus highest judge in the respective district or courthouse.

\*\* Portland probability level dropped to .066 for active judge test. Grand Rapids was only significant on all judge test, not significant for continuous service, and only one active judge so couldn't be compared on final test. Manhattan dropped to just above probability level at .054 for active judge test. Tampa was significant only for all judges, not significant for continuous service and active judge tests.

### Notes

- <sup>1</sup> Highlights from this study were posted as *Surprising Judge-to-Judge Variations Documented in Federal Sentencing*, Mar. 5, 2012, available at <http://trac.syr.edu/tracreports/judge/274/>. Figures included in this article are reprinted with permission from TRAC.
- <sup>2</sup> Ryan W. Scott, *Inter-Judge Sentencing Disparity After Booker: A First Look*, 63 Stan L. Rev. 1 (2010); Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity*, 90 J. Crim. L. & Criminology 239 (1999).
- <sup>3</sup> It is the nature of empirical research that study conclusions invariably rest in part on some untested assumptions. Central to the empirical results reported here are the assumptions that cases are assigned on a random basis to judges at the same courthouse and that prosecutors are unable to steer cases to particular judges within that same random assignment pool.
- <sup>4</sup> The development of this database, along with the associated analytical tools and judge reports, are described in *About the Data* at <http://trac.syr.edu/judges/aboutData.html>. For more information about obtaining reports or data extracts from the underlying database files, see *Obtaining Access* at <http://trac.syr.edu/judges/obtainingAccess.html>.
- <sup>5</sup> *About the Data*, *supra* note 4.
- <sup>6</sup> We note the relative absence of published research that systematically examines this issue, or confirms that federal criminal cases are being randomly assigned.
- <sup>7</sup> This does not imply that judges are the only individuals who exert influence over sentence length. Federal prosecutors influence sentence outcome in many ways, including through plea negotiations and of course in the decision on the specific charges brought. Unless federal prosecutors can steer cases to specific judges, or the judge assigned to a case influences which federal prosecutor is assigned to litigate a matter, random assignment also controls for any potential independent effects of the federal prosecutor assigned on inter-judge differences in sentencing outcome. The judge before whom a prosecutor appears can, of course, influence that prosecutor's decisions. This provides an indirect mechanism through which a judge can exercise influence over the sentence and is properly part of any measured inter-judge differences in sentence length.
- <sup>8</sup> Excluding judges who had decided few cases had little impact on case coverage in the study since they accounted for less than 1 percent of sentencing decisions. Their exclusion reduced the

number of sentences analyzed from 374,779 cases originally to 371,826. Since we examined the number of decisions a judge made within a specific district, many of these exclusions weren't actually of the judge per se but of the decisions rendered by that judge when they occasionally served as a visiting judge in another district. Also excluded from this analysis were three district court judges who are not Article III judges but were appointed to serve in U.S. territories.

- <sup>9</sup> If a judge had chambers at more than one courthouse in a district, the judge was assigned to a composite "office" that included each courthouse. For example, a judge having chambers in both Atlanta and Gainesville would be identified in the "Atlanta-Gainesville" office. The judge would then be compared only to other judges with the identical composite set of chambers, e.g., Atlanta-Gainesville.
- <sup>10</sup> We also ran analyses restricting comparisons to courthouses with at least three judges, and then those with at least four judges, who met our criteria to see if this had any influence on our findings. Other than reducing the number of offices compared; it did not affect the general pattern of our results.
- <sup>11</sup> For example, Payne (1997) noted: "Based on the distribution of the data, it is difficult to maintain the assumption that the prison terms are normally distributed." As a result, alternative significance tests were used, such as the Kruskal-Wallis test for differences in median sentences (A. Abigail Payne, *Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts*, 17 Int'l Rev. L. & Econ. 337 (1997)).
- <sup>12</sup> However, many of the differences were statistically significant, not only at less than the .05 level (less than 5 percent of the time, differences as large could occur by chance), but at less than the .00001 level or lower (less than one out of a 100,000 chance that differences as large could occur by chance). Such a minuscule probability suggests that in some courthouses these differences were quite large.
- <sup>13</sup> There were 672 judges that sentenced at least 40 drug offenders and 338 judges that sentenced at least 40 white-collar defendants and could therefore be compared. On average, judges compared had sentenced 154 drug offenders and 63 white-collar criminals. In total, these judges decided 103,750 drug cases and 21,149 white-collar cases. We used the Department of Justice's consolidated program categories to assign cases to crime categories in these comparisons. Although for reasons of space we limit our coverage to drug and white-collar comparisons, we did examine additional program

areas. Immigration offenses, although frequent, were primarily concentrated in a relatively small subset of districts—nearly three out of four occurred in the five districts along the southwest border. With weapons offenses, the next most frequent in occurrence in cases handled by district court judges, there was a rapidly diminishing number of judges handling 40 such cases that also severely reduced the number of courthouses that could be compared. We also used other case groupings, such as lead charge. In general, significant differences in median sentences were found among judges in many courthouses irrespective of the selection criteria used.

<sup>14</sup> St. Paul had only one active judge who had served during the entire period so could not be compared. St. Paul showed a significant sentencing difference when senior judges were included, and all cases, not just drug cases, were compared.

<sup>15</sup> Nebraska also presents an interesting example. Its busiest courthouse in Omaha showed statistically significant differences among judges in median sentences for their caseload as a whole. However, differences for drug sentences and for white-collar crimes were surprisingly small.

<sup>16</sup> See *Obtaining Access*, *supra* note 4.