Guest Editors’ Observations:
Using Data for Policymaking, Litigation, and Judging

Paul J. Hofer
&
William P. Adams

Guest Editor Paul J. Hofer is a Senior Research Associate in the Office of Policy Analysis at the United States Sentencing Commission.

Guest Editor William P. Adams is a Research Associate with the Urban Institute’s Justice Policy Center and an analyst for the Federal Justice Statistics Resource Center.

The views expressed are those of the guest editors and do not necessarily reflect a policy or position of the United States Sentencing Commission, the Urban Institute, or the Bureau of Justice Statistics.

A (Potentially) Dangerous Liaison
Policymakers, litigators, and judges, like most other Americans, have a love-hate relationship with data and statistics. Researchers working in public policy have heard the clichés and jokes more times than they can count: “You can say anything you want with statistics.” “Facts speak louder than statistics.” “There are three kinds of lies: 1) lies, 2) damn lies, and 3) statistics.” “If you laid all the statisticians in the world head to toe they still wouldn’t reach a conclusion.”

At the same time, policymakers and litigators are often eager to use data to bolster their own arguments, even if they remain skeptical of statistics. Compared to anecdotes or speculation, statistics hold the promise of impartiality and scientific objectivity. The progress made in fields like medicine by using data wisely reinforces the belief that public policy and legal decision making would also benefit from empirical research. Especially over the past few decades, the proliferation of data collection systems and computerized data processing have laid the groundwork for widespread use of federal criminal justice statistics in policymaking, litigation, and judging. But few would argue that the promise of “social science” has been realized. What accounts for the skepticism and the neglect?

In our experience, the integrity of the data should not be considered a problem. We have never seen anyone “cook the numbers” because they didn’t like a particular result. Sometimes categories get stretched, as illustrated by the discussion in this issue of data on federal “terrorism” prosecutions. And numbers used as work measurements, on which future funding may depend, should always be considered very carefully. And it is true that all statistics contain some measurement error, mistakes of classification or sampling. Part of the challenge for a good statistician is to determine the reliability of the data and provide reasonable estimates of the margin of error. But researchers working in public policy know that the data they report will be used in partisan debate and litigation. Great emphasis is placed on accuracy and impartiality. You simply cannot say anything you want with statistics. The numbers are either there or not.

The problem with statistics comes with their need for interpretation, especially as we try to use them to answer complicated questions of social policy. Like the world they describe, many data sets are highly complex and changing. To draw sound conclusions, interpreters must understand how the definitions of variables may have changed through the years, or how different samples were drawn. Many important policy questions are not readily addressed by data collected for administrative purposes. Criminal justice statistics are rarely collected in controlled conditions to test research hypotheses. Questions of causation, in particular—those vital questions about whether a policy change has had the desired effect, or whether a treatment program has worked—can almost never be answered in a straightforward way. With certain statistical tools of the trade some possibilities can be excluded and the field of plausible explanations can be narrowed. But a single, unqualified answer for many important questions may never be provided.

This need for qualification can be frustrating to policy advocates and lawyers who want to make the strongest possible case. The caveats of a careful statistician reinforce the cynical suspicion that there is no right answer, that the same numbers might support a different conclusion. Combating this cynicism, while insisting on the scientific integrity of their conclusions, is a key issue for researchers working with policymakers and litigators. Just because a single definitive answer cannot be provided does not mean that some conclusions aren’t better supported than others. And even a qualified and uncertain conclusion based on empirical evidence is usually more accurate than a policymaker’s hunch or a perception based on a few isolated examples. The proper comparison for a research result is not with perfect understanding but with the understanding we would have without the benefit of data.

Despite the potential for mistrust and misunderstanding, the marriage of policymaking, litigation, and judging with statistics and research is likely to continue and grow stronger. More and better data are becoming more readily available. These data will be, and should be, used with increasing frequency and skill. As noted by Professor and Federal Sentencing Reporter Editor Steve Chanenson, data and sentencing are not really an odd couple once their relationship is properly understood. This Issue is intended to inform readers of some
of the data concerning sentencing that are available today and to illustrate how it might be used with minimum confusion and maximum effect.

Finding and Interpreting Federal Sentencing Data
Before policymakers or litigators and sentencing data can marry they first have to meet. Happily there are several on-line services that connect potential data users with the information they need. For the old-fashioned, the Bureau of Justice Statistics (BJS) and the Sentencing Commission still publish print versions of their annual statistical reports. Every policymaker should spend an afternoon with the “bible” of criminal justice data, BJS’s SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS, which is updated every year and is also available online. It contains a wealth of information on topics such as public opinion about crime, police and courts, crime and victimization rates, and criminal justice expenditures and outcomes. The BJS-sponsored Federal Justice Statistics Program, which is operated by the Urban Institute and described in this issue by Bill Adams and Mark Motivans, collects and integrates data from different federal agencies at each stage of the criminal justice process. The program generates the annual reports COMPRENDUM OF FEDERAL JUSTICE STATISTICS and FEDERAL CRIMINAL CASE PROCESSING and it also generates reports on special topics, such as immigration offenses. All of these are available both in print and online.

The most detailed published information on federal crimes committed and the sentences imposed in any given year can be found in the Sentencing Commission’s SOURCEBOOK OF FEDERAL SENTENCING STATISTICS, which is also available both in print and online. How these data are collected and made available to the public is described in an article by Charles Loeffer of the Commission’s research staff. The Table of Contents of the SOURCEBOOK provides an explanation of the information contained in its 61 Tables and 14 Figures. In addition to data on offenses and offenders and the sentences imposed, the SOURCEBOOK contains special sections with more detailed information on drug, immigration, organizational offenses and a section on sentencing appeals.

These print sources can illuminate many issues but they often are inadequate for persons who need an answer to a specific question. The tables and graphs in the standard reports are designed to be of general interest but they may not “slice” the data in the way a user needs, or “cross-tabulate” two variables to provide the comparisons a user wants. For example, one can find in the Commission’s SOURCEBOOK the average prison sentence for most major offense types, but not necessarily the average prison sentence for offenders convicted under a specific statutory provision. Similarly, one can find a table in the SOURCEBOOK that provides the departure rates in various judicial circuits and a different table that provides the departure rates for different types of offenses. But if one is interested in comparing the departure rates in different circuits for different types of offenses, the tables in the SOURCEBOOK won’t help. They don’t cross-tabulate the data in that particular way, that is, they don’t break it down along the circuit and offense type variables simultaneously.

The online sources provide data and breakdowns that are not available in the printed reports. Many people are not aware, for example, that the Commission’s website contains circuit and district breakdowns that are not provided in the SOURCEBOOK OF FEDERAL SENTENCING STATISTICS. At http://www.ussc.gov/JUDPACK/JP2001.htm one can find each circuit and district compared to national data on the types of offenses sentenced, the mode of conviction, the types of sentences imposed, and the departure rate for each type of offense.

If these more detailed breakdowns are not sufficient, or if data from agencies in addition to the Sentencing Commission might be helpful, users should explore the Federal Justice Statistics Resource Center (FJSRC) at http://fjsrc.urban.org. The query system at the FJSRC site, described in this issue by Adams and Motivans, is the most useful tool available for persons not trained in computerized data analysis. Users can ask the system to generate tables on a wide range of issues and to break the information down in ways tailored to a particular question. For example, one can get information on the types of sentences imposed for offenders convicted under a particular statutory provision. Or one might cross tabulate sentence type with gender to see if probation is imposed more frequently on women than on men. The possibilities are endless. Some detailed information from pre-sentencing, sentencing, and post-sentencing stages is also provided at the district level on the “district statistics” page at the website, although district is not available for use in cross-tabulations in the query system due to concern that individual cases might too easily be identified.

Another source for federal criminal justice data is the Transactional Records Access Clearinghouse (TRAC), which is described in this issue by its founders David Burnham and Susan Long. TRAC can be found at http://trac.syr.edu. While including some sentencing information, TRAC’s emphasis is on the operation of the major federal law enforcement agencies. The site includes data on budgets, staffing, types of cases investigated and prosecuted, and a wide range of other information on federal investigating agencies. Because TRAC does not incorporate the Sentencing Commission’s datafiles, the information available on the nature of each offense and the sentence imposed is less extensive than the data available from the Commission or the FJSRC and may differ somewhat from these other sources.

TRAC offers some information for free, organized according to six major investigative agencies. One can learn, for example, the mean and median prison
sentence imposed on persons referred from the FBI or DEA or under various Department of Justice-identified program categories. The free site also provides the special topic reports for which TRAC is becoming well known, such as the recent report that suggested federal judges continue to sentence offenders disparately. The site also provides background information on the history and operations of federal investigative agencies.

TRAC also offers a variety of paid subscription services, which provide access to a wider selection of data and to additional analytical tools. For example, at the criminal section of the TRACfed subscription site, users can request trend data on a variety of measures, such as agency referrals, prosecutor declinations, convictions obtained, and average prison terms imposed. TRAC’s interactive analysis system allows users to request a variety of additional analyses, though we find its unique vocabulary — “going deeper,” “data slice,” “focus,” “rank” — to be somewhat off-putting. TRAC will conduct additional analyses for users at an additional charge. It seems likely that journalists will be intrigued by TRAC’s promise to help users “better understand the previously secret workings of the federal government.” The TRAC service is also unique in that users can perform analyses on individual judges or Assistant United States Attorneys. This relatively new development has been praised by some commentators and condemned by others. Burnham and Long suggest several ways that judge and prosecutor information might be useful to researchers and to Congress.

Mark Fazlollah, a reporter for the Philadelphia Inquirer, describes how using TRAC to investigate the performance of individual judges is an improvement over relying on the Almanac of the Federal Judiciary, which collects reviews of judges by private defense attorneys. This Almanac, which has been quoted over the years by journalists as an authoritative source, amounts to the subjective impressions of a non-representative sample of attorneys. Conceding that such anecdotal information is unreliable and potentially inaccurate, Fazlalloh turned to TRAC to confirm or refute assertions made about certain judges. He found little evidence to support the extreme comments found in the Almanac. Treating attorneys’ perceptions as “data” on which to draw conclusions is suspect and unfair, not to mention unscientific. Furthermore, it is unnecessary when statistics about judicial sentencing are available. However, Fazlollah rightly cautions that limitations of the TRAC data, such as missing or incomplete information, must also be carefully considered before reaching conclusions that may distort the record of a particular judge or prosecutor.

The Need for Complete Data and Methodological Rigor

Kevin Blackwell, a senior research associate with the U.S. Sentencing Commission, highlights some additional pitfalls of inappropriately using federal sentencing data, particularly the identities of individual judges. He points out that some analyses conducted by TRAC, such as its recent report on sentencing disparity, lack the rigor found in most scholarly explorations of the issue. Simple comparisons of the mean and median sentences imposed by different judges tell us little about whether judge-created disparity exists. Such comparisons ignore differences in offense and offender characteristics that judges are required to consider. The influence of these legally relevant factors must be adequately controlled, and researchers have developed statistical methods for doing so. Without use of these methods, reporting “disparities” between judges can create a false impression that the differences in average sentences found are unwarranted.

Researchers interested in more rigorous exploration of complicated questions must obtain the full datasets and perform the analyses themselves. Adequate training in statistics and research methods are required for such endeavors, however. The datasets can be obtained from several sources. The FJSRC files can be downloaded directly from their website. The Sentencing Commission makes most of its datafiles available through the Inter-University Consortium for Political and Social Research (ICPSR). Confidential information, including the identity of sentencing judges, is not included in these files and may be obtained only in the limited circumstances discussed further below.

Researchers also must ensure that they understand the definitions of offenses and how these definitions have changed over time. Nora Demleitner, a law professor and Federal Sentencing Reporter managing editor, describes how the apparent upward trend in the number of federal terrorism prosecutions may largely be an artifact of an expansion in the types of cases counted as terrorism by the Department of Justice. Shifting definitions always presents challenges to interpreting trends; this is why seasoned criminal justice statisticians cling to standard offense definitions whenever possible. Demleitner offers some explanations for why the definition of terrorism has expanded, including the budget and career incentives that U.S. Attorney’s Offices and investigatory agencies may have in appearing “tough on terrorism.” On the other hand, she cautions that alternative ways of dealing with terrorism suspects, such as extradition, rendition, or deportation, may mean that DOJ’s statistics might under-represent the full range of the government’s anti-terrorism efforts. Demleitner concludes that the data show that prosecutors are more likely than before to charge low-level offenses as terrorism-related. This appears supported by a recent report released by TRAC, which found very low median prison sentences imposed for terrorism-related convictions — for example, two weeks for “international terrorism.” Instead of indicating judicial leniency, this
finding suggests that the Department of Justice has an expansive definition of what constitutes a terrorism offense.

**Using Data Fairly and Effectively**

As discussed above, there can be tension between effective advocacy and objective analysis. In zealously arguing for a position, data might be misinterpreted or attended to selectively. Several papers in this Issue illustrate the tightrope-walk involved in effectively using data in policy advocacy and litigation without distorting the facts as you find them.

Bill Mercer, the United States Attorney for the District of Montana, makes a strong case that the original data collection and reporting requirements for departures from the sentencing guidelines were proven inadequate. His argument helped convince Congress to enact the reporting provisions found in the PROTECT Act of 2003, which became effective on April 30. Few could argue with Mercer’s claim that better data will facilitate evaluation of the guidelines and help determine whether departure rates are excessive. Nor can one dispute the data showing that the downward departure rate (for reasons other than substantial assistance to the government) has increased over the last twelve years. Identifying the causes of this increase, however, is a trickier business. As Mercer shows, the rates vary significantly among circuits and districts. Discovering whether the increases vary among individual judges within the same district could also be illuminating. The relative contributions of “fast track” programs, the Supreme Court’s decision in *Koon v. United States*, and other policy changes that may have eroded the strictures of the guidelines must all be considered.

Mercer’s assertion that the increasing downward departure rate signals that the Sentencing Reform Act’s goal of minimizing unwarranted disparity is endangered rests partly on legal and policy judgments in addition to the data, and also on empirical assumptions extending beyond the available data. Certainly the effects on disparity of “hidden departures” in the form of unjustified charge dismissals or fact and sentence bargains, or disparate use of USSG §5K1.1 or Fed. R. Crim. P. 35(b) motions must also be considered. But Mercer’s conclusions are plausible and additional research, using the improved data made available by the PROTECT Act, will be needed to ensure that departures are not underminding the Sentencing Reform Act’s goal of minimizing unwarranted disparity.

Barbara O’Connor, the First Assistant Federal Defender in the District of Vermont, provides several examples of how data might be creatively used to refute misconceptions and challenge conventional wisdom. In an era when plea agreements are critical and prosecutors have been bound by less rigid guidelines than judges, data may be needed to educate an inexperienced Assistant U.S. Attorney on what the “going rate” for a particular type of crime really is, or to inform the court of how often a particular guideline adjustment is really applied. Certainly, the traditional guidelines concept of the “heartland” case—and its associated ideas of “typical,” “exceptional,” or “rare” circumstances—appear to make statistics relevant to deciding when a guideline should be applied or when a departure would be appropriate. O’Connor argues that the recent enactment of the PROTECT Act makes the use of data in advocacy even more important than it was before.

**Using Data to Evaluate Judges**

Data on the sentencing behavior of individual judges have not routinely been publicly available. A longstanding agreement between the Sentencing Commission and the Administrative Office of the United States Courts, on whose cooperation the Commission has depended for all its data, prohibits the Commission from releasing datasets that include the identities of individual judges except to qualified researchers. These researchers must be affiliated with recognized research institutions and engaged in Commission-approved work. They must enter into cooperative agreements with the Commission and promise not to publically divulge confidential information, including the identities of judges.

In recent months, two new mechanisms for the release of judge-specific information have appeared. The first is the TRAC website noted above, which allows users to compare individual judges and Assistant United States Attorneys on the variables it receives from its various sources. The second are provisions of the PROTECT Act that require the Commission to make sentencing documents, which include the identity of sentencing judges, available to the House and Senate Judiciary Committees upon their request. In addition, the Commission’s electronic datasets, including judge-specific information, must be made available to the Department of Justice upon their request. No statutory provision has been made to release this judge-identifying information to the defense bar or researchers. Thus, there now appear to be three mechanisms by which judge-specific information may be made available to certain users outside the judicial branch.

Some journalists and academics have long argued that the identities of sentencing judges should be publicly released along with data on the sentences they have imposed. They point out that this information is available to anyone with the resources to gather it in open court or from public records in individual courthouses. At the federal level, it is only the national compilation of the information in the Commission’s datasets that is kept confidential. Those who believe the best solution is to “let the sun shine in” and loosen data release policies will find support from the experience of Pennsylvania, as reported in the article by Mark H.
Bergstrom and Joseph Sabino Mistick of the Pennsylvania Commission on Sentencing. The same tensions between judicial independence and accountability heard now at the federal level were voiced in Pennsylvania in the late 1990s. Over the objections of many members of the judiciary, the Commission adopted a liberalized data release policy in 1999. The Commission held public hearings on the policy and supplemented it with training for judges and the media, with opportunities for judicial review and correction of data, and other safeguards. The result has been improved data collection, an increase in the amount of research conducted, and heightened accountability for the Commission. These and other ancillary benefits have, in the opinion of these authors, outweighed the occasional misuse of the data in newspaper stories that over-simplify particular judges’ sentencing decisions.

Would full public release of judge-identifying sentencing data be preferable to the three mechanisms currently available? Initial experience with reports using the partial sentencing data available through TRAC provide some reasons for optimism but many reasons for concern, as shown in the articles by Fazlollah and Blackwell. Experience is lacking with release to Congress and the Department of Justice under the provisions of the PROTECT Act. It remains unclear what the Judiciary Committees and DOJ will do with the data they may request in the future. The recent episode concerning Judge Rosenbaum of the District of Minnesota, whose sentencing records were requested, under threat of subpoena, after he delivered controversial testimony concerning sentencing policy before the House Judiciary Committee, along with the recent formation of the House Working Group On Judicial Accountability to monitor the decisions of federal judges, naturally has aroused concern that sentencing records will be used in an attempt to intimidate judges and threaten the independence of the judiciary.

So what, if anything, should be done? One’s answer seems to depend largely on how responsibly one believes the data will be used. There may be ways to reduce the chances for misuse while making more data available. For example, confidential judge codes instead of actual judge names could permit some types of research that have not been possible, while making misuse—such as misleading and sensationalistic “Judicial Report Cards” that have appeared in newspapers in some states—less likely. It certainly behooves all the parties with access to judge identifying information—including TRAC, the Commission, the Judicial Conference, and Congress and the DOJ—to prove their responsibility by using the data conscientiously and making it available to others appropriately.

**Making Rational Sentencing Possible: The Promise of Data and Research**

The great promise that data and research hold for judging and policymaking once the complex issues of data interpretation and access are resolved are illustrated by three papers in this Issue. Cyrus Tata and Neil Hutton, Co-Directors of the Centre for Sentencing Research at Strathclyde University in Glasgow, Scotland, write about the Sentencing Information System (SIS) now in place in the High Court of Scotland. This system is an alternative to sentencing guidelines that holds great appeal for many judges. Instead of prescribing sentencing policy, SIS simply describes sentencing patterns found in cases similar to a case the user inputs to the system. It also provides details of similar individual cases in a narrative and qualitative manner. The system is voluntary and the criteria for defining the similarity of cases have been left largely to the judges themselves. Time will tell if Scottish judges, unlike their Canadian counterparts, actually use and maintain the system, or if its popularity is based more on the wish to avoid more restrictive sentencing guidelines than on a real interest in the sentencing practices of other judges.

Judge Michael Marcus of the Circuit Court of Multnomah County, Oregon, takes a different approach to sentencing data. He is concerned not merely with past sentencing practices, but with what sentences have been effective for reducing recidivism and controlling crime. The sentencing information system he envisions is a warehouse of the accumulated knowledge of criminal justice research. Judge Marcus’s review of the “state of the art” in risk prediction and treatment program evaluation is a vision of rational sentencing informed by data. His article is an indictment of the status quo and a plea for data collection and research aimed at crime control instead of “just deserts.” The computerized sentencing information system Judge Marcus has begun to develop aims to make available the recidivism rates of offenders, similar to the offender being sentenced, who were sent to various alternative dispositions in the past. For example, how does the recidivism rate of an offender like this who was given probation compare to one who was given shock incarceration? The logic of the system is impeccable, but cautious researchers will question whether the data can support the detailed inferences he hopes to make. Quantifying the effects of different treatment programs for different types of offenders is a very tricky business. As Judge Marcus recognizes, “sentencing support tools are roughly where aviation was at the Wright Brothers’ first flights at Kitty Hawk.” But as a vision of the way ahead his work merits serious attention.

Finally, the article by Professor Alfred Blumstein of Carnegie Mellon University considers the use of data for legislative policy making. He reminds us of the need to re-examine sentencing policy in light of new empirical information as it becomes available. Performing this exercise on the differential treatment of crack and powder cocaine under current sentencing statutes and guidelines, he concludes that the 100 to 1 quantity ratio
is in urgent need of reconsideration. One can only hope that policymakers will soon see the benefits of relying on research instead of rhetoric and act to improve the fairness and effectiveness of federal sentencing policy.

Will the marriage of data and sentencing be successful? The articles in this Issue lend support to the optimists and to the skeptics. The stakes are high in terms of fairness and public safety. The investment of time and money in collecting data, building institutions like the Sentencing Commission, the FJSRC, and TRAC, and in educating policymakers, litigants, and judges in how to use the data has already been great. The marriage of data and sentencing requires still more work, but if successful it could help us all live more happily ever after.

Notes
3 See U. S. Sentencing Commission, DOWNWARD DEPARTURES FROM THE FEDERAL SENTENCING GUIDELINES (October 2003).
4 The Commission’s data release policy can be found at http://www.ussc.gov/general/publicacces.pdf