An Overview of U.S. Sentencing Commission Data

The Commission's data collection and research mission

In keeping with its mandate to serve as a "clearing-house and information center for the collection, preparation, and dissemination of information on Federal sentencing practices," the U.S. Sentencing Commission has compiled data files on federal sentencing practices since soon after the inception of guideline sentencing at the federal level. The Commission has also collected data files for a limited number of years prior to the advent of guideline sentencing. These files are used by the Commission and outside researchers to study a range of subjects related to criminal sentencing in the U.S. courts.

The Commission regularly uses data compiled by its Office of Monitoring and analyzed by its Office of Policy Analysis to inform the decision-making process on amendments to the sentencing guidelines. For example, prison population projections, which are required by statute for any guideline change that might require an expansion in correctional capacity, would be difficult to make without the detailed data contained in the datasets. The Commission also receives frequent requests for information on sentencing from the Congress, other interested parties, the general public, and members of the judicial branch.

Outside researchers, many of them affiliated with universities have also made good use of the data compiled by the Commission. A range of articles on the sentencing guidelines themselves or changes in criminal or judicial behavior have been published by authors who have found the datasets compiled by the Commission to be a valuable source of information on federal sentencing practices.

The Commission's mandate also includes establishing sentencing policies and practices that "reflect, to the extent practicable, advancement in knowledge of human behavior as it relates to the criminal justice process," and developing "means of measuring the degree to which sentencing, penal, and correctional practices are effective in meeting the purposes of sentencing as set forth in section 3553(a)(2) of title 18, United States Code." It is therefore crucial for the Commission to compile this data and for researchers both within and outside the Commission to have access to this information in order to better understand the practice of criminal sentencing in the federal courts.

Sources of data on federal sentencing

The Commission compiles datasets from criminal cases involving convictions for federal felonies and Class A misdemeanors sentenced under the U.S. sentencing guidelines. Since the enactment of sentencing guidelines, every federal court is required to send to the Commission copies of court documents for guideline cases including the indictment(s), the judgment of conviction, the presentence report, the statement of reasons, any written plea agreement, and any revocation of probation or supervised release.

Once case documents are received from the federal courts, Commission staff members enter a variety of sentencing information into the data collection system. From this system three separate datasets are created for each fiscal year: the monitoring dataset, the appeals dataset, and the organizational defendant dataset.

The monitoring dataset is available from fiscal year 1988 onwards and includes sentence information from the judgment of conviction order, departure information from the statement of reasons, as well as background demographic and guideline application information from the pre-sentence report. In this dataset, each record represents an individual adult criminal defendant. Due to limitations of the early monitoring datasets, the Commission staff generally recommends using datasets from fiscal year 1991 onwards.

The appeals dataset is available from fiscal year 1993 onwards. It contains information on criminal cases in which any appeal was filed including the district where the appeal was filed, the circuit where the appeal was filed, the date of appeal, the date of opinion, legal issues before the court, and the court's disposition.

The organizational defendant dataset is available from fiscal year 1993 onwards. It contains information on organizations sentenced under Chapter 8 of the sentencing guidelines. Included in this dataset is information on the organizational structure, organiza-
tional size, economic viability, offense of conviction, mode of adjudication, sanctions imposed, and application of sentencing guidelines.

On occasion the Commission compiles more specialized datasets in order to examine particular subjects not adequately covered by the monitoring datasets. Examples of such special study datasets include the Augmented FPSSIS dataset, the Mandatory Minimum Study Sample Dataset, and the 1991 Evaluation Datasets. Whenever the Commission compiles such a dataset, it generally issues a publicly accessible report and places a copy of the underlying dataset at the Inter-University Consortium for Political and Social Research (ICPSR).

A more detailed list of the Commission’s datasets is available in the annual Publications and Resources Guide from the Commission or the Commission’s website.

In addition, each dataset is accompanied by a comprehensive codebook to the variables contained within that dataset.

**Commission Policies Concerning Data Access and Use**

The Commission’s policy regarding public access to data on federal sentencing collected by the Commission was published in a Federal Register notice on December 13, 1989. This policy, which is still in effect, states that “the Commission will create a comprehensive dataset on federal sentencing practices under the Sentencing Guidelines that will be made available for public use through the Inter-University Consortium for Political and Social Research.” Furthermore, the Commission’s policy states that these datasets will be updated with new cases and the information contained in them will be “consistent with its [the Commission’s] agreement with the Administrative Office of the U.S. Courts regarding the confidentiality of certain documents.”

While much of the information found in the documents provided to the Commission is a matter of public record, not all of the information is public. Observers at a sentencing hearing in any courtroom can learn which guidelines are applied in any sentencing and whether the offender has any criminal history. But other information contained in the pre-sentence report for the judge’s use and included in the datasets, such as a defendant’s substantial assistance to the Government, is generally not made public even at the sentencing hearing. Including this information in the publicly released datasets, but deleting the identifying information, permits some types of research that would not otherwise be possible.

In accordance with the terms of the memorandum of understanding between the Commission and the Administrative Office of U.S. Courts the datasets publicly released through the ICPSR and the Federal Justice Statistics Resource Center (FJSRC) do not contain identifiers for particular judges or defendants. Researchers may still request access to this and other data in accordance with the procedures outlined in the Commission’s published data access policy, but should be aware that successful requests are rare.

The Commission’s agreement with the Administrative Office also means, in combination with security and financial considerations, that no “source documents” (i.e. pre-sentencing reports, judgment and conviction orders, written plea agreements) will be made available to the public either. With these exceptions, the Commission makes available for public use all other regularly collected information on federal sentencing.

**Accessing Commission Compiled Data on Federal Sentencing**

The Commission makes the data it collects and compiles available to the public through three primary mechanisms: dissemination of datasets, publication of statistical sourcebooks, and web-based documents. Each of these formats offer particular advantages to the researcher.

By far the most detailed data on federal sentencing are the datasets themselves, available for no charge from the Inter-University Consortium for Political and Social Research (ICPSR) located at the University of Michigan and from the Federal Justice Statistics Resource Center (FJSRC) maintained by the Urban Institute. However, the effective use of these datasets requires at least a basic knowledge of statistics and data management, as well as a basic understanding of the federal sentencing guidelines themselves. The monitoring datasets contain a wealth of sentencing information, allowing researchers to reconstruct precisely which guideline provisions were applied in each case. Thus, all of the important legally relevant offense and offender characteristics can be reconstructed by determining which guidelines and which specific offense characteristics were applied. Each dataset is accompanied by a detailed codebook, which describes the variables contained in that dataset in greater detail.

The annual Sourcebook of Federal Sentencing Statistics is available in a PDF format on the Commission’s website and in paper upon request from the Commission’s public affairs office. The sourcebook is a rich source of information on many aspects of federal sentencing including average sentences by offense, use of guideline adjustments, frequency of out-of-range sentences, demographic characteristics for individuals sentenced, and number of appeals initiated. In general, it serves as an excellent starting point for those interested in federal sentencing. The Sourcebook replaced the research section of the Commission’s Annual Report in 1996, and was expanded and updated. Information is generally provided in a standardized format for each year, and short-term trends are included for several important guideline factors.
Finally, the Commission publishes regional break-downs of certain statistics on its website. These packets provide certain district-by-district and circuit-by-circuit comparisons that are not included in the annual Sourcebook.

Other Sources of Data on Federal Sentencing
In addition to the data files deposited with the ICPSR and the FJSRC, and the summary data available from the Commission itself, several organizations make searchable information on federal sentencing available to the general public. These include the Federal Justice Statistics Resource Center and the Transactional Records Access Clearinghouse, both of which allow specific data queries for little or no fee and are discussed in more detail elsewhere in this issue. An alternative starting point is The Sourcebook of Criminal Justice Statistics, which contains considerable information on every stage of the state and federal criminal justice systems, including summary data on federal sentencing practices. It should be kept in mind that these alternatives access sites derive a majority of their information on federal sentencing from Commission supplied datasets. The researcher may prefer to use one of these organizations for convenience, but should be aware that most detailed national data on federal sentencing is derived from a single source.

Special Data Requests
On occasion the Commission receives requests for specific analyses of sentencing data. Generally speaking, these requests are handled as expeditiously as the Commission’s limited time and staff resources allow. When the requested information can be found more quickly in a publication such as the annual Sourcebook or some other source, the Commission directs the requester accordingly. Since the Commission is a small federal agency, researchers are often referred to the datasets available from the ICPSR and FJSRC. If the request is time-sensitive and the requester is a party to ongoing litigation, he or she is encouraged to have the presiding judge or probation officer make the request to the Commission. The Commission’s experience has been that many litigators are not familiar with the queries that may be made through the Federal Justice Statistics Resource Center website. This resource allows instantaneous results even for persons who do not otherwise have access to the Commission’s raw data or do not have statistical analysis capability.

Using Commission Compiled Federal Sentencing Data
Consulting published Commission reports and other relevant literature on federal sentencing is always a good starting point before undertaking a research project using federal sentencing data. All of the Commission’s reports are available either on the Commission’s website or upon request from the Commission itself. In addition, many researchers have found it useful to contact the Commission’s research staff directly with preliminary questions about working with the data on federal sentencing.

The primary datasets used by researchers and practitioners are the annual monitoring datasets, which contain information on criminal cases received by the Commission from the U.S. district courts for each fiscal year. Accompanying each data file is a codebook, which indicates the meaning of each variable and variable value within the dataset. An additional resource, available upon request from the Commission, is The Most Frequently Asked Questions About USSC Datafiles. This document contains helpful advice on understanding the structure of the datasets and the relationship between variables within the datasets.

Constitutional challenges to the guidelines, efforts to standardize document coding schemes, as well as normal implementation delays, led to a large number of missing cases in the early years of Commission compiled data. For these reasons, the earliest dataset that the Commission staff generally recommends using is the monitoring database from fiscal year 1991.

Statutory changes and guideline amendments happen with some frequency, and for that reason, it is important that the researcher be aware of any such changes. A good way of doing so is to look at the latest Guidelines Manual for guideline amendments bearing on guideline cases of interest. If there have been significant amendments, the researcher may find it necessary to consult earlier guideline manuals in order to determine what version of a guideline was in effect during the period of interest. Several of the Guidelines Manuals are available on the Commission’s website and the rest are available upon request from the Commission itself. In addition, a separate variable in the dataset indicates which edition of the Guidelines Manual was used in the sentencing of a particular case.

Finally, it is worth mentioning that administrative documents create an administrative database. Although the monitoring database is extremely “clean” and comprehensive by the standards of most administrative databases, expectations concerning its use should be realistic. The underlying documents which form the basis for the datasets are not specifically designed as research coding instruments, nor are thousands of individuals who fill out these documents trained as research document coders. For these reasons, researchers may find that variables they would have coded are not present in the dataset, or that the utility of a particular variable is unclear to them. These are routine concerns for anyone using a dataset not specifically designed for his or her use and it is a testament to the quality of the underlying documents and coding process that so much research can be done using this database.
Commission’s overall policies with regard to the public release of identifying information.

Notes
8. 54 Fed. Reg. 51279.
11. The provisions of the PROTECT Act, Pub. L. 108-21, that require the Commission to release certain information upon request to the Congress, and that require the Commission to provide electronic datasets containing judge identifiers to the Department of Justice, have not changed the Commission’s overall policies with regard to the public release of identifying information.
17. http://www.albany.edu/sourcebook