
*Evaluating Data on Corporate Sentencing:
How Reliable are the U.S. Sentencing
Commission's Data?*
(forthcoming in the FEDERAL SENTENCING REPORTER)

Cindy R. Alexander, Jennifer Arlen, and Mark Cohen

USC Olin Research Paper No. 01-7



RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

*This paper can be downloaded without charge from the Social Science Research Network
electronic library at http://papers.ssrn.com/paper.taf?abstract_id=269356*

**EVALUATING DATA ON CORPORATE SENTENCING
HOW RELIABLE ARE THE U.S. SENTENCING COMMISSION'S DATA?**

Cindy R. Alexander
Economic Analysis Group
U.S. Department of Justice
Washington, DC 20530
alexancr@usdoj.gov

Jennifer Arlen
University of Southern California Law School
Los Angeles, CA 90089
jarlen@law.usc.edu

and

Mark A. Cohen
Owen Graduate School of Management
Vanderbilt University
Nashville, TN 37203
Mark.cohen@owen.vanderbilt.edu

April 2001
revised

* Cindy R. Alexander: Economic Analysis Group, U.S. Department of Justice (Antitrust Division); Jennifer Arlen: University of Southern California Law School and California Institute of Technology (Visiting Professor); and Mark Cohen: Owen Graduate School of Management, Vanderbilt University. Work on this project was initiated while Alexander was Victor Kramer Fellow at the University of Chicago Law School. Jennifer Arlen would like to thank the National Science Foundation (grant no. SAR 973081) and the University of Southern California Law School for financial support. Mark Cohen acknowledges the financial support of the Dean's Fund for Summer Research, Owen Graduate School of Management. The authors gratefully acknowledge helpful comments from and discussion with members of the Commission staff at various stages of this research, and excellent research assistance from Mina Kim and Sandeep Satwalekar. The views expressed in this paper are authors' alone and do not necessarily represent the views of either the U.S. Department of Justice or the National Science Foundation.

ABSTRACT

During a recent study of how 1991 federal sentencing guidelines have affected the penalties that federal courts impose on public corporations, we performed an independent evaluation of the quality of the data on corporate sanctions that the U.S. Sentencing Commission releases to the public. Our initial findings led us to use other, independently-compiled data for our own research. This paper presents the main findings of our evaluation, which focused on the quality of the Commission's 1988-1996 (ICPSR) data on public corporations. First, the Commission's post-Guidelines data on penalties for public corporations appear to be incomplete and non-representative of the underlying case population. For example, the ICPSR post-1991 data appear to exclude a disproportionate number of large fines imposed on public corporations. No similar difficulties in the ICPSR pre-Guidelines (1988-1989) data were found. Shortfalls in the post-Guidelines data on other kinds of defendants, such as individuals, appear to be less marked. Second, the Commission's data are missing variables that may explain a substantial part of the case-by-case variation that occurs in sentencing. The data reveal little about the harm caused by the offense, which is often estimated in court papers. Also missing is information about the identity of the sentencing judge and about the identity of the corporation being sentenced. We review the history of the Commission's efforts to collect data on federal sentencing, highlighting institutional constraints and other factors that appear relevant to the difficulties we have found in the data that the Commission releases to the public.

The Commission collects and analyzes data on guideline sentences to support its varied activities. As authorized by Congress, the Commission's numerous research responsibilities include: (1) the establishment of a research and development program to serve as a clearinghouse and information center for the collection, preparation, and dissemination of information on federal sentencing practices; (2) the publication of data concerning the sentencing process; (3) the systematic collection and dissemination of information concerning sentences actually imposed and the relationship of such sentences to the factors set forth in section 3553(a) of title 18, United States Code; and (4) the systematic collection and dissemination of information regarding the effectiveness of sentences imposed (28 U.S.C. § 995(a)).

The Sentencing Commission maintains a comprehensive, computerized data collection system which forms the basis for its clearinghouse of federal sentencing information and which, in large part, drives the agency's research mission. Pursuant to its authority under 28 U.S.C. §§ 994(w) and 995(a)(8), and after discussions with the Judicial Conference Committee on Criminal Law and the Administrative Office of the U.S. Courts, the Commission requested that each probation office in each judicial district submit ... documents on every offender sentenced under the guidelines ...

– U.S. Sentencing Commission 1999. *Sourcebook of Federal Sentencing Statistics* (Fourth Edition).

INTRODUCTION

At the federal level, Congress entrusts the task of providing the public with comprehensive criminal sentencing data to the U.S. Sentencing Commission (hereinafter, the “Commission”), whose primary responsibility is to regulate federal sentencing practice. Little has been written, however, about the quality of the data that the Commission makes available to the public or uses itself for policy analysis. The Commission is charged with maintaining a complete database on federal sentencing practice. To date, researchers have largely assumed this goal has been achieved, relying on Commission data without independently verifying its quality. This reliance may have been misplaced.

During a recent empirical study of the 1991 U.S. Sentencing Guidelines for Organizations (hereinafter, the “Guidelines”) and their effect on penalty levels,¹ we independently investigated the

¹ For our complete analysis, see Cindy R. Alexander, Jennifer Arlen and Mark A. Cohen, "Regulating Corporate

quality of the data on federal corporate sentencing that the Commission makes available to the public through the Inter-University Consortium for Political and Social Research at the University of Michigan [hereinafter, ICPSR data] and other channels.² The results of our investigation raise concerns about the thoroughness of the data the Commission has compiled in the post-Guidelines era, and the adequacy of the information reported.

First, a substantial number of cases appear to be missing from the Commission's ICPSR data on organizational sanctions. Indeed, the data do not appear to be representative of the underlying case population. The post-Guidelines data appear to be missing a disproportionate number of large fines. This raises concerns about statistical bias.

Second, we found the Commission's ICPSR data to be missing variables critical to analysis of sentencing practice. For example, the post-Guidelines ICPSR data reveal little about offense severity, even though — as recognized in the Guidelines themselves — the appropriate sentence depends on the severity of the crime.³ Nor do the data reveal the judge's identity, even though there is evidence to

Criminal Sanctions: Federal Guidelines and the Sentencing of Public Firms," 42 *Journal of Law and Economics* 393 (1999) [hereinafter, Alexander, Arlen & Cohen, JLE]. For a review and discussion of findings, see Cindy R. Alexander, Jennifer H. Arlen and Mark A. Cohen, "The Effect of Federal Sentencing Guidelines for Public Corporations," 12 *Federal Sentencing Reporter* 20 (1999) [hereinafter, Alexander, Arlen & Cohen, FSR].

² These ICPSR data can be located at www.icpsr.umich.edu, by (1) typing in the study number, which is "9513," and searching words in the category, "in the study no." and then (2) clicking on "downloads" and following instructions.

³ For a recent discussion of optimal criminal sentences in relation to crime severity, see Jennifer Arlen and Reinier Kraakman, "Controlling Corporate Misconduct: An Analysis of Corporate Liability Regimes," 72 *NYU Law Review* 687 (1997). On how the optimal criminal sanction may be affected by the presence of other, non-criminal sanctions, see Cindy R. Alexander, "On the Nature of the Reputational Penalty for Corporate Crime: Evidence," 42 *Journal of Law & Economics* 489 (1999).

suggest that the identity and background of the judge affect sentencing practice.⁴ Finally, the Commission's ICPSR data do not include the name of the offender, corporate or individual. Without the identity of the offender, it is virtually impossible to combine the Commission's sentencing data with other data that can usefully shed light on sources of disparity that may persist among corporate criminal sanctions. These factors include the corporation's size and line of business and, most importantly, what other sanctions were imposed on the corporation for the same offense, including private civil liability, which the Commission's ICPSR data do not report.

One of our purposes in writing this article is to inform and caution researchers and others about some of these problems in the Commission's data on organizational sanctions in order to clarify their limitations. While the data remain a useful resource for sentencing research, it is realistic to be concerned about drawing inappropriate conclusions about sentencing trends from these data. For example, using the Commission's ICPSR data it appears that the median fine is lower for public corporations under the Guidelines than pre-Guidelines. Yet more complete, independently-compiled data suggest that fines are indeed higher under the Guidelines. It is our hope that drawing attention to these and other difficulties will highlight the value of providing support for Commission efforts to compile data for its own uses and for dissemination to outside sources — both in terms of funding and through the relaxation of institutional restrictions on data access.

For other readers, this article illuminates the process by which federal sentencing data are compiled, and the factors by which the quality of the data is assessed. A special National Academy of

⁴ See *infra* note 26.

Sciences (NAS) Panel on Sentencing Research in 1983 concluded a lengthy study by offering advice on how future studies of sentencing data should be conducted.⁵ The Panel stressed the importance of ensuring that sentencing data include, first, a complete, or at least representative, sample of cases from the relevant offender population and, second, information about all relevant factors. In explaining our experience with the Commission's data, we highlight the factors that have been relevant to our own research, on the sentencing of corporations. Similar issues arise in the data on sentencing of individuals, although those data do not seem to be as incomplete as the organizational sanctions data, as we explain.

Finally, and more generally, we seek to underscore the importance of investing in data as an important foundation for public policy, especially in sentencing. Without appropriate data, is it difficult to evaluate the real consequences of proposed changes in sentencing practice, or to investigate the practical validity of expressed concerns about their fairness or effectiveness. The development of sentencing policy — through Guidelines or otherwise — has in the past been guided by a clear understanding on the part of policymakers of existing practice: for example, of stated rules and the sentences actually implemented, and of the relationship between sentencing and wrongdoing. This requires good data.

Data, of course, come at a cost and it might seem reasonable to defer some of this expense now that public scrutiny of Commission efforts has waned, relative to what it was in the 1980's, when the Commission wrote its first Guidelines. Yet while deferring the cost of maintaining data can be an

⁵ See Alfred Blumstein et al., eds., *Research on Sentencing: The Search for Reform*, National Academy Press, Washington, D.C. (1983).

attractive short-term strategy, like any deferred maintenance plan, it can be risky. Demand for data to resolve Guidelines-related issues can arise suddenly. And misguided policy recommendations can emerge from analyses based on poorly maintained or incomplete data. Thus, the quality of the data remains important, even while the public profile of efforts to develop and maintain good data may shift.

Fortunately, our experience suggests that the Commission's institutional commitment to maintaining the quality of its data is fairly strong. This does not mean there are no challenges to overcome. Indeed, this article identifies serious problems in the data, some of which persist, and suggests areas for improvement. Yet the discussions we have had with Commission staff in preparing this article indicate that steps have been taken to correct some of the problems we have found, and that the staff in fact discovered some of these difficulties independently at the time of our own investigation, as we explain.

I. HISTORY OF COMMISSION DATA COLLECTION EFFORTS

For data, the Commission relies almost entirely on the cooperation of the district courts under a letter agreement with the Administrative Office of the U.S. Courts (hereinafter "Administrative Office"), dated June 22, 1988.⁶ The courts send the Commission paper copies of documents on relevant cases, such as the master docket, official indictment, pre-sentence investigation report, plea agreement, and judgment of conviction order. From these documents, the Commission extracts and codes information into electronic databases, which are analyzed internally and sent (after modification) to the

⁶ United States Sentencing Commission 54 *Federal Register* 51279-01 (1989) ("The cooperation of the Administrative Office in the collection of the data is essential to the Commission's ability to carry out its statutory mandate..."). For a recent discussion of Commission data collection procedures, see for example U.S. Sentencing Commission, *1998 Annual Report*, Chapter 5 at 35 ("Statutory Requirements: Data Collection").

Inter-University Consortium for Political and Social Research (ICPSR) for posting over its public-access website.⁷

The ICPSR data on organizational sanctions have been compiled in distinct phases, using different methods and focusing on different variables (with similar variables sometimes defined differently). The first phase focused on cases sentenced during 1984-1987.⁸ This was followed by efforts to obtain data for 1988 and, soon thereafter, for cases sentenced between January 1, 1989 and June 30, 1990. The Commission used the 1988-90 data as background in writing its 1991 Guidelines,⁹ which were the first data to be posted at the ICPSR website.

In compiling “pre-Guidelines” data, the Commission staff periodically verified the completeness of data from the courts by comparing them with the contents of the Administrative Office’s independent criminal Masterfile. The staff also made follow-up calls to the courts helped to ensure the data were complete.

After promulgating the 1991 Guidelines, the Commission shifted from research mode to monitoring mode. It made major adjustments to its data collection strategy at that time.¹⁰ First, it

⁷ United States Sentencing Commission, 54 *Federal Register* 26132-02 (1989).

⁸ See Mark A. Cohen, Chih-Chin Ho, Edward D. Jones, III., and Laura M. Schleich, "Organizations as Defendants in Federal Court: A Preliminary Analysis of Prosecutions, Convictions and Sanctions, 1984-1987," 10 *Whittier Law Review* 103 (1988) and Mark A. Cohen, "Corporate Crime and Punishment: A Study of Social Harm and Sentencing Practice in the Federal Courts, 1984-1987," 26 *American Criminal Law Review* 605 (1989). A public use version of the 1984-87 data (absent any identifying information that would reveal the name of convicted offenders) are available from Cohen on request.

⁹ For recent discussion of the role that sentencing data played in the early development of federal sentencing Guidelines, see Marc Miller & Ron Wright, “Your Cheatin’ Heart(land): The Long Search for Administrative Sentencing Justice,” 2 *Buffalo Criminal Law Review* 723 (1999).

¹⁰ On the limited scope of the post-Guidelines’ data collection effort, see U.S. Sentencing Commission, 1993, “Data Collection Description” in “Organizations Convicted in Federal Criminal Courts, 1987-1993 (United States): Part 6: Organizational Defendants Data, 1987-1993,” ICPSR, publisher (“...despite the fact that prior Commission research

began to rely more heavily on the initiative of the courts for data, and it discontinued staff follow-up efforts to ensure all relevant case files were being sent.

Second, the Commission changed the type of cases for which it sought data. Previously, data were obtained on all federal convictions. Post-Guidelines, data were obtained only for cases explicitly governed by the Guidelines. This dramatically reduced the number of case records sought because some federal crimes by organizations are not governed by the Guidelines, and the Guidelines are binding only on crimes occurring on or after November 1, 1991.

Third, the Commission changed the type of information it sought post-Guidelines, focusing more on information necessary to implement the Guidelines, and less on information to evaluate their effectiveness. The post-Guidelines data exclude information on the dollar value of the loss caused by crime, previously included. They exclude the identities of the judge and the organizational defendant. Finally, the data do not reveal whether the organization had an effective compliance program, reported the offense or otherwise cooperated with law enforcement officials, even though these factors are relevant to the Commission's stated goals.¹¹

II. MISSING CASES

To investigate the quality of the Commission's ICPSR data, we undertook to compile data

indicates that there were approximately 300 organizational defendants sentenced annually, few organizational defendants have been sentenced pursuant to Chapter Eight to date.”).

¹¹ The Commission reports whether a convicted organization (e.g., corporation) received a lower penalty (i.e., mitigation) for monitoring, cooperating or reporting. Yet, given the Guidelines' structure, one cannot confidently discern whether an organization actually monitored, investigated or reported from data on whether it received mitigation because an organization may not receive mitigation for one of these activities even if it engaged in it.

independently of what is available through the Commission and published by the ICPSR.¹² Our exhaustive effort focused on public corporations convicted of federal crimes, 1988-96. Our sources included a keyword search of *Lexis/Nexis* and *Westlaw* (month-by-month, state-by-state), the *Wall Street Journal Index*, SEC 10-K filings, the Department of Justice Alert data base, the *Corporate Crime Reporter*, and the Commission's data on organizational offenses.

The contrast between the Commission's ICPSR data and what we were able to obtain from public sources is striking. The post-Guidelines ICPSR data show only 13 cases in which public corporations received criminal sentences under the Chapter Eight Guidelines, 1992-1996. By contrast, we found 34 cases in which public corporations appear to have received sentences under the fine provisions of the mandatory Guidelines in that period — more than double the cases in the ICPSR data. Even in the 1996 data alone, we found about twice as many such cases in our independent data as in the ICPSR data. If we also count cases that appear not to have been Guideline-constrained, the number of public corporations with federal criminal sentences in the 1992-1996 independent data set grows to 142 — about ten times what is in the ICPSR data.

This shortfall in the Commission's ICPSR data appears to be traceable to the shift the Commission made from research mode to monitoring mode around the time it issued the 1991 Guidelines. Consistent with this, cases are missing from the post-Guidelines ICPSR data, yet there is no evidence of a shortfall in the pre-Guidelines data. Rather, there are about as many cases in the pre-Guidelines ICPSR data as in the pre-Guidelines independent data — 49 public corporations sentenced

¹² For detailed discussion of our method and underlying rationale, see Alexander, Arlen & Cohen, *JLE*, *supra* note 1, and relatedly Cindy R. Alexander and Mark A. Cohen, "Why do corporations become criminals? Ownership, Hidden Actions, and Crime as an Agency Cost," 5 *Journal of Corporate Finance* 1 (1999).

in 1988-89 according to the ICPSR data versus 46 according to the independent data (i.e., 24.5 versus 23 per year), as Table 1 shows.

TABLE 1
NUMBER OF FEDERAL CONVICTIONS OF PUBLIC CORPORATIONS, BY DATA SOURCE

A. Pre-Guidelines (All)			
Year	Independent Data	ICPSR Data	% difference
88	19	19	0.0%
89	27	30	11.1%
90	29	17*	n.a.
B. Post-Guidelines (Guidelines cases only)			
Year	Independent Data	ICPSR Data	% difference
92	0	0	0.0%
93	3	1	-66.7%
94	11	3	-72.7%
95	10	5	-50.0%
96	10	4	-60.0%
97	n.a.	6	n.a.
98	n.a.	2*	n.a.

*: ICPSR 1990 and 1998 data are only available for a half-year (1/1 - 6/30);
n.a : Independently-compiled data do not cover the 1997-98 period.

A. *Which cases are missing?*

The evidence of missing cases is most striking for public corporations. The post-Guidelines ICPSR data on Guidelines offenses contain less than half as many cases as our post-Guidelines independent data on Guidelines offenses for the same period. Similarly, post-Guidelines ICPSR data contain about one-third the convictions one would expect to find had underlying pre-Guidelines case levels and case mixes continued into the post-Guidelines period.

The shortfall appears less acute for organizations other than public corporations, e.g., private corporations, non-profits and municipalities. The post-Guidelines ICPSR data contain about 80 percent as many cases as pre-Guidelines sentencing patterns lead us to expect, although we have no

independent data on these other types of organizations.

The smallest case deficit appears to arise in the data on individual offenders, based on the Commission's own research. The Commission in 1997 undertook a "data completeness" project, finding that those data appear to comprise at best about 90 percent of the case population.¹³

No one seems to know why some of the Commission's post-Guidelines data are incomplete. One possibility arises from the Commission's reported practice of obtaining documents exclusively through judicial district probation offices.¹⁴ If probation officers were less closely involved in handling documents for corporate offenders, this could explain some of the shortfall in the organizational data. Other possibilities might arise by careful further examination of the process by which case records are obtained.

To be sure, the Commission has taken steps to improve the completeness of its organizational sanctions data, starting with the FY 1999 data. It has made similar efforts with its data on individual offenders annually since 1997. Both projects involve comparing data received from the courts with what is in the Administrative Office Masterfile.

This substantial progress does not to our knowledge change, however, the Commission's post-Guidelines practice of excluding information on cases to which the federal Guidelines do not expressly apply. This unfortunately excludes a substantial number of federal criminal cases from the data. For example, federal guidelines have not yet been written on the sentencing of organizations for

¹³ See United States Sentencing Commission, *1997 Annual Report to Congress*, at 34, 48 and Appendix A.

¹⁴ See "Chapter 5: Research — Data Collection" in U.S. Sentencing Commission, *1998 Annual Report on Commission procedures for compiling raw sentencing data*.

environmental crimes, so the data exclude environmental cases. Thus, even if the data were complete by the Commission's current standards, they would reveal nothing about the sentencing of those cases — nor about how sentencing under the express control of the Guidelines compares with sentencing practice elsewhere in the federal criminal system. This limits their practical usefulness.

B. *A non-representative sample*

Of greater concern than missing cases, the post-Guidelines ICPSR data on organizational sanctions appear to be non-representative of the underlying case population. Research employing the Commission's post-Guidelines ICPSR data may thus find lower post-Guidelines fines and case counts for public corporations than is in fact the case — particularly when compared with pre-Guidelines fines and case counts from that source — because the post-Guidelines data appear to exclude a disproportionately large number of cases in which big fines were imposed.

For example, relying on the post-Guidelines ICPSR data alone, one might conclude that the median criminal fine for public corporations was *lower* under the Guidelines than it had been previously: \$109 thousand post-Guidelines (1992-96; n = 13) versus \$265 thousand pre-Guidelines (1988-89; n = 49). Indeed, relying on those data, a recent article concluded that corporate fines are not significantly greater post-Guidelines than pre-Guidelines.¹⁵

¹⁵ Relying on ICPSR data, Jeffrey S. Parker and Raymond A. Atkins “find no statistically significant change in the level or structure of corporate monetary penalties imposed under the guidelines during 1992-95 as compared with baseline data taken from pre-Guidelines cases sentenced in 1988.” Jeffrey S. Parker and Raymond A. Atkins, 1999, “Did the Corporate Criminal Sentencing Guidelines Matter? Some Preliminary Empirical Observations,” 42 *Journal of Law & Economics* 423 (1999), discussed in Alexander, Arlen & Cohen, *JLE*, *supra* note 1 and Alexander, Arlen & Cohen, *FSR*, *supra* note 1, at note 22.

In fact, this impression from the ICPSR data is misleading. In contrast with the ICPSR data, our independent post-Guidelines data suggest public corporations received about \$486.4 million in criminal fines in 1996. Most of this — 94 percent — appears to have been imposed under the Guidelines, based on the type of crime and when it occurred. This not only exceeds the \$105.7 million the Commission’s ICPSR data report as paid by *public corporations* in 1996, but also exceeds the amount they report as paid by *all* organizations in Guidelines cases that year.

TABLE 2

AVERAGE CRIMINAL FINES FOR PUBLIC CORPORATIONS (000\$)*

Regime (Era)	ICPSR Data		Independent Data	
	Mean	Median	Mean	Median
Pre-Guidelines (1988-89)^	\$834	\$265	\$1,081	\$633
Ch. 8 Guidelines (1992-96)	\$8,859	\$109	\$19,051	\$3,095

* All fines are expressed in thousands of 1996 dollars (using CPI deflator);

^ ICPSR reports half-year 1990 only, making comparison difficult. Neither reports 1991 cases.

The greater criminal fine volume in our independent data does not simply arise from the greater number of cases in those data. The ICPSR data appear disproportionately to exclude larger sentences. Our data indicate that the median fine under the Guidelines was \$3.1 million in 1992-96 (n = 34) — almost five times the pre-Guidelines median of \$633 thousand (1988-89; n=45), as Table 2 shows. Moreover, while the post-Guidelines ICPSR data show that only two organizations received a criminal fine greater than \$25 million under the Guidelines in 1992-96, our independent data include five cases in which public corporations alone received such fines (based on reports of pleas or sentences in 1994, 1995 and 1996). Analyzing these data, we report elsewhere that criminal fines for public corporations

tend to be significantly higher under the Guidelines than before¹⁶ — in contrast to what the ICPSR data suggest. See Table 3.

TABLE 3
DISTRIBUTION OF CRIMINAL FINES FOR PUBLIC CORPORATIONS

Regime (Era)	ICPSR Data			Independent Data^^		
	Below \$1MM*	\$1MM or more	Total	Below \$1MM*	\$1MM or more	Total
Pre-Guidelines (1988-89)^	39 (80%)	10 (20%)	49 (100%)	31 (69%)	14 (31%)	45 (100%)
Ch. 8 Guidelines (1992-96)	8 (62%)	5 (38%)	13 (100%)	13 (38%)	21 (62%)	34 (100%)

* All fines are expressed in 1996 dollars (measured using CPI deflator).

^ ICPSR reports half-year 1990 only, making comparison difficult. Neither reports 1991 cases.

^^ Counts exclude cases in the data for which fine information is missing.

The evidence thus suggests that the post-Guidelines data on organizational sanctions available through ICPSR constitute a non-representative sample of organizational Guidelines cases. This conclusion — that some of the post-Guidelines ICPSR data are non-representative — appears also to apply to the data on *individual* offenders, based on the Commission’s own research.¹⁷ This is striking

¹⁶ See note 1, *supra*.

¹⁷ See U.S. Sentencing Commission, *supra* note 14. For example, although the Commission noted that the distribution of offense types was largely the same, it found three exceptions: (1) drug trafficking convictions were more likely to be included in the original data, (2) immigration cases were less likely to be included, and (3) “other” mostly minor offenses were less likely to be included. Thus, researchers are to be cautioned about drawing inferences about changes in the proportion of these types of cases over time. For example, a common “stylized fact” about federal criminal law is that drug cases are taking up a significant proportion of the caseload. While certainly true, it is interesting to note that in 1989-90, the Commission reported 47.7% of Federal Guidelines sentences were for drug offenses, compared to 39.2% in fiscal year 1997. It is not known how much of this “decrease” is due to a relative lessening of the drug caseload versus the fact that in earlier years, the Commission data were more likely to include drug offenses than others. Similarly, care must be taken when comparing immigration cases, which accounted for 6.3% of the 1989-90 cases, but 13.7% of those in fiscal year 1997. We do not know if more immigration cases are being prosecuted or if more are being reported by district courts. Although not fully analyzed, the Commission also noted that gaps in reporting were not equally distributed by district court. Thus, geographic trends also must be analyzed with care.

in light of the 1983 NAS Panel report, recommending the use of data that comprise either a full set of sentenced offenders or a representative sample from the population of interest.¹⁸ If the data are *not* representative, there is a risk that any inferences — and hence policy implications — from the data might be wrong.¹⁹

C. *Are there other explanations?*

The conclusion that substantial numbers of cases are missing from the ICPSR data persists even after considering several alternative explanations of our findings. For example, we noticed some of the cases in our post-Guidelines data were not formally sentenced as of the last year of the sample. This is because those public corporations had publicly committed to a penalty under a plea agreement but had not yet been formally sentenced. Since the ICPSR data exclude cases not yet formally sentenced, this could account for some of the disparity. Yet the same disparity does not arise in the pre-Guidelines ICPSR data. Nor does the lag between plea and sentencing appear to account for the fact that some large fines in the independent 1992-96 data are not found in the ICPSR data set. The dates on some of those fines seem too early for them not to have been imposed formally by the last year of the data.

Alternatively, we considered the possibility that the Commission might have delineated “public corporations” more narrowly than we have done in our independent data, since we included penalties

¹⁸ See Blumstein et al., *supra* note 5.

¹⁹ For example, a researcher employing the Commission’s ICPSR data set might well conclude that the Commission had not achieved its stated goal of increasing corporate sanctions simply because the post-Guidelines data tend to under-report cases in which large fines have been imposed. This could mislead policymakers into seeking changes to the Guidelines that would implement the original goal, even though that goal of increasing corporate criminal fines had in fact already been achieved.

imposed on the private subsidiaries of public corporations. This does not appear to explain the disparity, however.

III. MISSING VARIABLES

The ICPSR data also do not contain all relevant information about the cases they report.²⁰ This omission limits the usefulness of the data for making contemporaneous comparisons among sentences, not just for analyzing how sentences change over time. The difficulty is most noticeable in the post-Guidelines ICPSR data.

After issuing the Guidelines, the Commission revised its list of variables coded from court documents to add over a dozen variables capturing the mechanics of how courts apply the Guidelines. Unfortunately, it excluded other important variables, some of which were included previously. The problem of omitted variables is always a challenge for sentencing research. It can cause researchers to attribute outcomes erroneously to one factor that are really due to another, omitted factor. We have identified several variables whose omission from ICPSR data make it more difficult to analyze the Guidelines' effect on organizational sanctions: (a) monetary loss, (b) name of offender, and (c) name of sentencing judge.

A. *Offense Loss*

To make sense of sentencing trends, one must have a benchmark for comparison. Otherwise, it

²⁰ For a list of factors relevant to the analysis of sentence disparity, see 18 USC 3553(a), the relevance of which is highlighted in 28 USC 995(a)(15) (on "Powers of the Commission").

is difficult to tell whether higher fines, for example, reflect more severe sentences or more severe crimes (or both).²¹ Data on offense loss can act as such a benchmark. The pre-Guidelines ICPSR data indeed report “loss” in two ways — “loss for all related conduct” and “loss for only offenses resulting in conviction.” In contrast, the post-Guidelines data do not separately or consistently report loss.²² We had expected to find a loss variable in the data because it affects the fine under virtually any theory of sentencing.²³

The practical usefulness of the Commission’s data would be enhanced markedly if the Commission would code from court documents *all* factors that have historically been found related to sentence levels, such as loss, and not just those that play an explicit role in the Guidelines.

B. *Identity of the Offender*

The NAS Panel’s challenge to consider “all relevant factors” when analyzing sentence disparity is daunting, particularly for organizational offenses. For example, a corporation’s size, financial condition, ownership structure and market conditions all are related to the occurrence of crime and its consequences. They are thereby related to penalties and their disparity.²⁴

²¹ For discussion of alternative measures of loss, see Alexander, Arlen and Cohen, *JLE*, *supra* note 1, at 411-412.

²² To be sure, the post-Guidelines ICPSR data report estimates of “loss or gain” for some cases. Yet there is no information on which of the two numbers is being reported — victim’s loss or offender’s gain. This makes the information of limited use for producing estimates of crime magnitude, even where it appears.

²³ Indeed, the Guidelines instruct judges to set the base fine at the greater of (i) the Guideline-prescribed fine based on offense severity, (ii) the organization’s pecuniary gain from the offense, or (iii) the pecuniary loss from the offense to the extent the loss was caused intentionally, knowingly or recklessly. U.S. Sentencing Commission, *Guidelines Manual* (November 1, 1991), § 8C2.4.

²⁴ On the relation between ownership structure and corporate conduct, see Cindy R. Alexander and Mark A. Cohen, *JCF*, *supra* note 12. For evidence on financial and market conditions in relation to crime, see Cindy R. Alexander and

Most of this is not in court documents, yet numerous data bases exist from which researchers could readily obtain the information *if* they knew the convicted corporation's identity. With this information, the Commission staff can itself merge variables from outside sources into its data. The Commission also can enable outside researchers to do so by releasing two pieces of information: (1) the name of the organizational defendant and (2) a description of the misconduct.

Unfortunately, the Commission's June 1988 agreement with the Administrative Office prevents it from releasing offenders' names to the public. The ban is on revealing information "that will identify an individual defendant or other person identified in the sentencing information." It is not clear to us how this furthers the public interest. Convictions are a matter of public record. Even if there might be arguments for not releasing convicted *individuals'* names in a public data set, we know of no persuasive arguments for withholding names of convicted *organizations*.²⁵ This decade-old agreement substantially limits the usefulness of the Commission's ICPSR data in shedding light on the sources of sentence disparity and in otherwise illuminating the Guidelines' effectiveness in furthering the goals of the Sentencing Reform Act.

C. *Sentencing judges as relevant factors*

Finally, we turn to the sentencing judge. Certainly, the name of the judge would appear on

Mark A. Cohen, "New Evidence on the Origins of Corporate Crime," 17 *Managerial & Decision Economics* 421 (1996).

²⁵ There appears to be some statutory authority for withholding names of defendants from data sets in the criminal justice system, 42 USC 3789g(a). In the case of individual defendants, it appears the main concern is over the privacy rights of the individual whose prior criminal history might be revealed. Corporations are normally considered "persons" in the eyes of the law. Yet the rationale for extending this treatment to organizations, especially public corporations, is unclear.

many researchers' lists of "relevant factors." Research has shown that judge-related factors can matter in some situations.²⁶ Indeed, a primary reason for establishing the Commission was Congressional concern over inter-judge disparities in sentencing.

Yet notwithstanding the importance of understanding how differences among judges can affect sentencing, researchers cannot easily determine the name of — nor even a unique identifying code for — the sentencing judge associated with each individual case. This is true even though the relevant information is publicly available on a case by case basis from each district court, although not in a centrally-accessible format.

The difficulty is that the Administrative Office has been reluctant to release the name (or code) of the sentencing judge in any of its data sets. In responding to the Sentencing Commission's first request for baseline sentencing data in 1986, the Administrative Office deleted the judge code from its data before forwarding those data to the Commission, despite the Commission staff's explicit request for this information to help it understand the nature of pre-Guidelines sentence disparity, as it had been directed to do.²⁷ Judge identifiers were later restored and provided to the Commission under the terms

²⁶ For empirical evidence on how identity of individual judge may affect case outcomes, see, e.g., Mark A. Cohen, "Explaining Judicial Behavior or What's 'Unconstitutional' about the Sentencing Commission?" 7 *Journal of Law, Economics and Organization* 183 (1991); Gregory C. Sisk, Michael Heise and Andrew P. Morriss, "Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning," 73 *NYU Law Review* 1377 (1998); A. Abigail Payne, "Does Inter-Judge Disparity Really Matter? An Analysis of the Effects of Sentencing Reforms in Three Federal District Courts," 17 *International Review of Law and Economics* 337 (1997); James M. Anderson, Jeffrey R. Kling, and Kate Stith, "Measuring Interjudge Sentencing Disparity: Before and after the Federal Sentencing Guidelines," 42 *Journal of Law & Economics* 271 (1999), and Joel Waldfogel, "Aggregate Inter-Judge Disparity in Federal Sentencing: Evidence from Three Districts," 4 *Federal Sentencing Reporter* 151 (1991). See also, Mark A. Cohen, "The Motives of Judges: Empirical Evidence from Antitrust Sentencing," 12 *International Review of Law and Economics* 13 (1992).

²⁷ This information is based on 1986 discussions between Mark A. Cohen and the Chairman of the Commission.

of the letter of agreement prohibiting release of any individual's name the public.²⁸ Thus, they are not in the public ICPSR data.

Adding the name of the judge to the ICPSR data would facilitate empirical investigation of whether judge-related disparity, documented in past research, persists under the Guidelines. Adding the name of the judge would be sufficient because it could be linked to other sources of data on factors previously suggested as related to judges' decisions, such as political party, demographic profile, and prospects for appointment to a higher bench. Even without the name of the judge, however, the addition to the ICPSR data of a unique identifying code for each judge would usefully support progress in this area. For example, it would facilitate analysis of whether residual disparity were occurring within a judge's sentences or across judges.

IV. CONCLUSION

Following promulgation of the Guidelines, the Sentencing Commission has seen its mission evolve to include a wide array of responsibilities. Among these are maintaining sentencing data for public use and the conduct of sentencing research.

As researchers and "consumers" of the Sentencing Commission data, we are impressed by the scope of the data collected and the assistance the Commission has provided to researchers.

Nonetheless, we have identified two difficulties undermining the reliability and usefulness of the published data. First, some cases are missing altogether, and this raises questions about the

²⁸ United States Sentencing Commission, 1989, *supra* note 6 at 51282 ("No information that will identify an individual defendant or other person identified in the sentencing information will be disclosed to persons or entities outside the Commission...").

representativeness of the data. Second, information that could shed light on sentence disparity and help formulate sensible policy is not currently coded or made publicly available.

The Commission is a relatively new organization whose place among federal agencies is still evolving. It has faced considerable external pressure since its inception, and considerable commissioner and staff turnover. Unlike the Bureau of Justice Statistics (BJS) and other Federal agencies that have been around for decades, there is not yet a large research community engaged in using the Commission data and thereby helping ensure their quality and completeness.

Now that there is once again a full slate of Commissioners with an active research and policy agenda, we can expect sentencing data to take on an increasingly important role in policy discussions. This is an opportune time for the Commission to revisit the importance of maintaining truly comprehensive federal sentencing data, and for others to renew their budgetary and institutional support for its efforts.