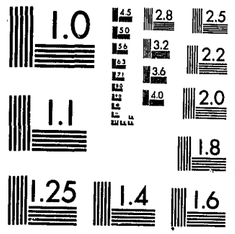
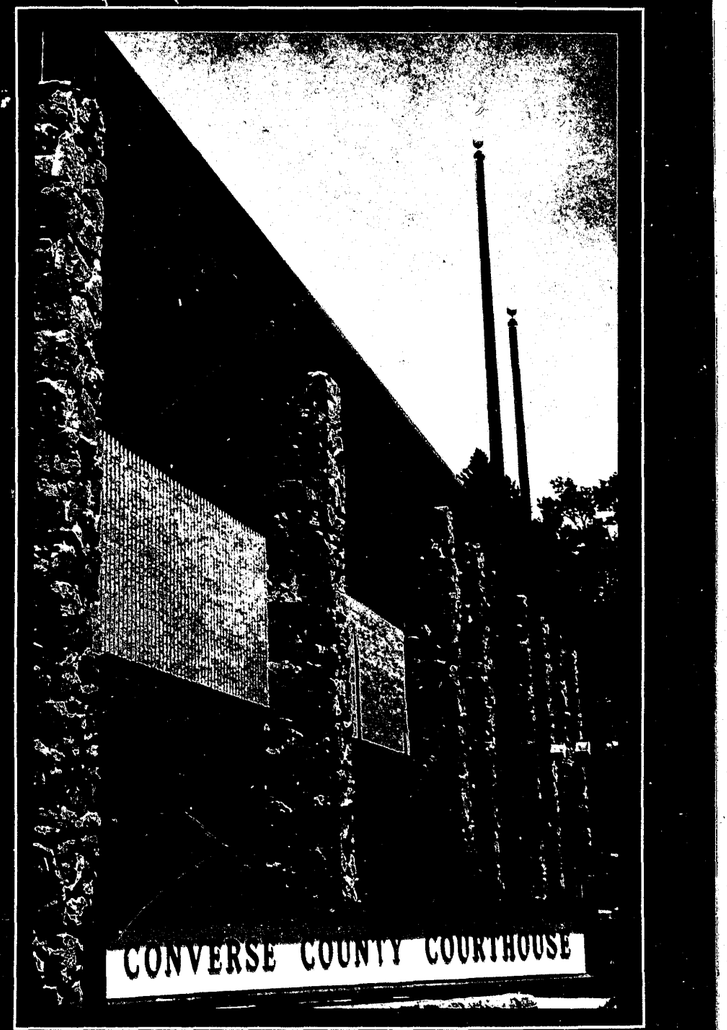
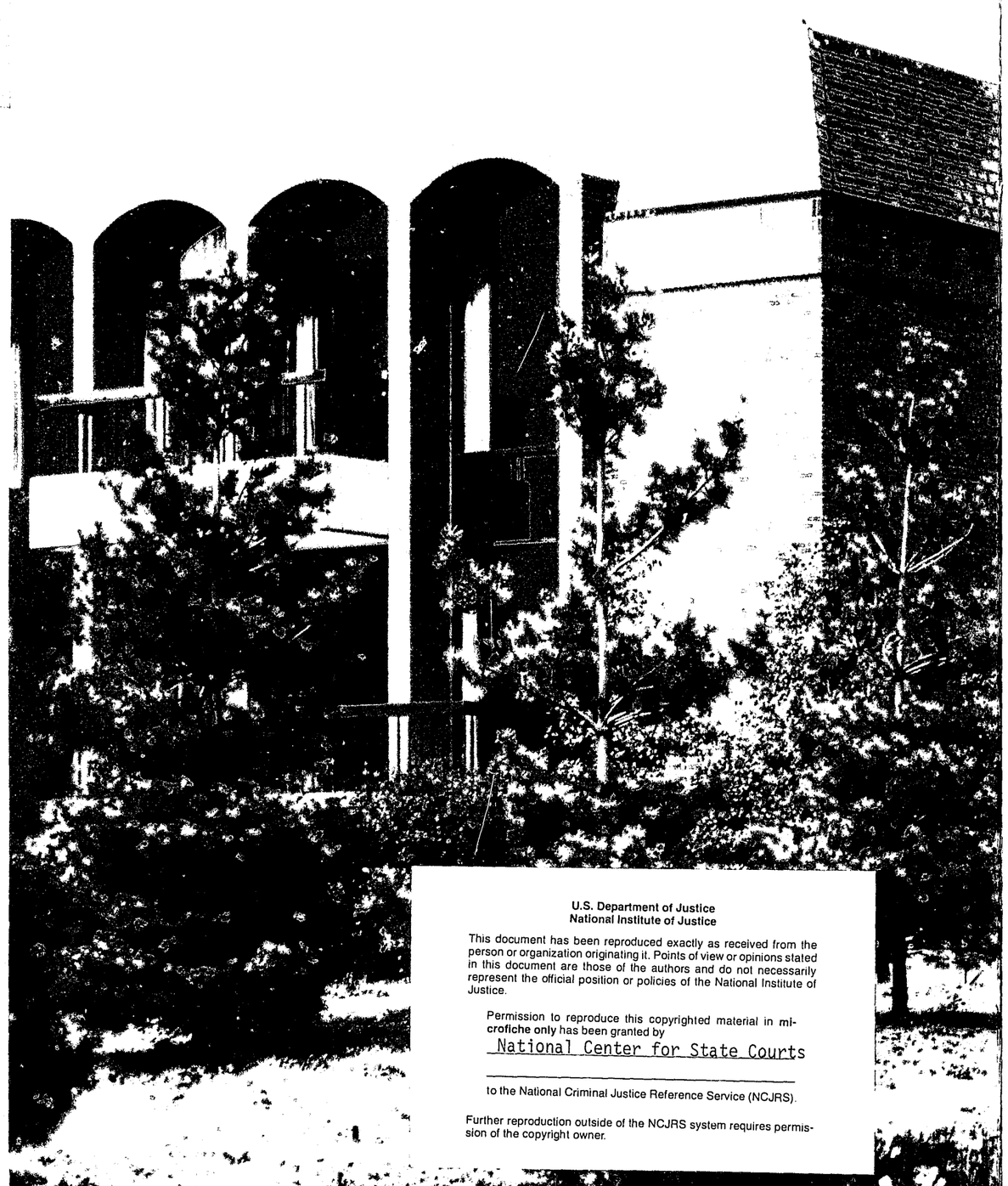


Published by the National Center for State Courts





National Center for State Courts 1971

State Court Journal

Summer 1982 JRB Volume 6, Number 3

Special Section on Judicial Immunity

Features

Departments

On the Cover

The Converse County Courthouse in Douglas, Wyoming, was designed by Henry Therkildsen and built by Julien Construction Company. Completed in 1976, the two-story structure houses all county offices, including the county and district courts and the jail. Photo by Mark Cash.

Illustrations on pages 3 and 32 by Marilou Gilley.

Sentencing by mathematics

A National Center study evaluates several early attempts to develop and implement sentencing guidelines.

By William D. Rich, L. Paul Sutton, Todd R. Clear, and Michael J. Saks

During the past decade there has been tremendous pressure to reform sentencing. Along with the abandonment of the rehabilitative ideal has come a demand for limits on the broad sentencing discretion that characterized the era of indeterminate sentencing. In response to public concern about disparity in sentences, several state and local jurisdictions have attempted to structure judicial sentencing discretion by means of empirically based sentencing guidelines. These guidelines have been developed on the basis of statistical models of prior judicial sentencing behavior, and they purport to represent the average sentences in various kinds of cases. They are intended not to alter the overall severity of sentences but to reduce unjustified differences by encouraging judges to gravitate toward the average sentence for a particular kind of case, as defined by several factors pertaining to the offense and the offender. Compliance with the guidelines is voluntary, and provisions are made for judges to depart from the sentence recommended by the guidelines by giving reasons in justification of the departures.

Inasmuch as sentencing guidelines, either judicially adopted or administratively promulgated, are emerging as the principal alternative to legislatively fixed or presumptive sentences, it seemed important in our study to gather empirical evidence about the efficacy of the guidelines reforms that have been implemented thus far. Our report addressed itself to three general issues: (1) the logical and methodological problems of creating sentencing guidelines by means of statistical models of past judicial sentencing behavior; (2) the impact of sentencing guidelines on judicial sentencing decisions; and (3) the relationship between sentencing guidelines and the exercise of discretion by prosecution and defense.

This evaluation focuses on sentencing guidelines that have been adopted by judiciaries at the local or county level. The four courts that served as sites for this study were the Denver District Court in Denver, Colorado; the Philadelphia Court of Common Pleas in Philadelphia, Pennsylvania; the Cook County Circuit Court in Chicago, Illinois; and the Essex County Court in Newark, New Jersey. The Denver, Chicago, and Newark courts participated in a project to develop and implement sentencing guidelines, which was funded by the Law Enforcement Assistance Administration (LEAA) and conducted by the Criminal Justice Research Center (CJRC); the Philadelphia court conducted its own project without funding by LEAA but with some informal assistance from the CJRC.

Our research was based on (1) reanalyses of the case record data used in the Denver sentencing guidelines feasibility research and the Chicago and Newark sentencing guidelines implementation research; (2) analyses of case record data that we collected as part of this project in Denver and Philadelphia; and (3) interviews that we conducted in Denver, Chicago, and Philadelphia. Thus, our examination of guidelines' empirical basis was limited primarily to Denver, Chicago, and Newark, with primary emphasis on Denver, whereas our assessment of the impact of sentencing guidelines on sentences and the exercise of discretion by prosecution and defense was limited exclusively to Denver, Chicago, and Philadelphia, with primary emphasis on Denver and Philadelphia.

The research design of the quantitative assessment of the impact of the Denver and Philadelphia guidelines was that of a case study. Each court was considered a case with various characteristics, such as an incarceration rate, an average sentence length, and so forth. These courts were selected because they were the only courts suitable and available for study, not because they were necessarily representative of courts in general. Therefore, we can make no statistical inferences about all courts. Nevertheless, the application of substantive knowledge and sound judgment permit the formulation of conclusions that are likely to obtain in other jurisdictions.

This study sought, in part, to measure the impact of sentencing guidelines on several characteristics of the Denver and Philadelphia courts. In order to do so, we compared the characteristics of each court before the implementation of guidelines with its characteristics afterwards. Strictly speaking, this is a nonexperimental, single-group, pretest-posttest design. For some purposes we measured the characteristics of each court at several points in time before and after the

Prepared under Grant Number 78-NI-AX-0140 from the National Institute of Justice, U.S. Department of Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice or the National Center for State Courts.

adoption of sentencing guidelines. To this extent, the design resembles that of an interrupted time series, but none of the sophisticated statistical techniques often associated with that design were employed.

The limitations of the research design were such that no single statistic should be taken as conclusive evidence of the proposition it tends to support. Where, on the other hand, multiple measures and our informal descriptive observations pointed to the same conclusion, our confidence in that conclusion is augmented.

The data for the before-after comparisons and time series were collected from individual case files in the Denver and Philadelphia courts. These data included large amounts of information about the offense, the offender's background and criminal history, and the processing of the case. The Denver data set included all convictions for which charges were filed during the period beginning 18 months before the implementation of guidelines (in November 1976) and ending 24 months after the implementation date. These data therefore constituted the entire population of cases in the Denver court rather than a sample in the usual sense of the term. There were 3,605 convictions in the Denver data set, of which 1,208 belong to the preguidelines subset and 2,397 belong to the postguidelines subset. The Philadelphia data set consisted of a 10-percent random sample of cases in which sentences were imposed during the 26 months that preceded the implementation of sentencing guidelines (on March 5, 1979) and a 45-percent sample of cases in which sentences were imposed during the six months following the implementation date. The postguidelines sampling ratio was greater than the preguidelines sampling ratio in order to provide a sufficient number of cases in which to study compliance.

Personal interviews were conducted with judges, prosecutors, defense attorneys, and, to a lesser extent, court administrators, probation officials, and prisoners. Most of these interviews were semistructured, and included both closed- and open-ended questions. A total of 169 interviews were conducted. These were supplemented by self-administered questionnaires given to respondents upon completion of the interview.

Although a great deal of the report is concerned with the presentation and interpretation of statistical analysis, much of what we learned came from the perceptive comments of the persons we interviewed. We have employed journalistic as well as social scientific methods in an effort to learn as much as possible about our subject.

The Empirical Basis of Sentencing Guidelines

The most salient feature of the sentencing guidelines examined in this study is their putative empirical basis. Their claims to legitimacy and judicial support depend upon their empirical validity. The first part of the study addressed two interrelated questions. First, are the sentencing guidelines empirically valid? Second, is the process by which the guidelines were created an appropriate way to formulate sentencing policy? We addressed these questions by reviewing the capabilities and limitations of statistical modeling techniques, examining the procedures used by the guidelines' developers, reanalyzing the data that were used to create the Denver, Chicago, and Newark guidelines, and analyzing some of our own data collected for the impact study.

The guidelines' source of legitimacy and acceptance is the assertion that they embody the implicit sentencing policies which judges collectively have followed in the past. The reasoning behind this assertion can be exemplified in its simplest terms as follows: if we observe that offenders with one prior felony conviction received sentences that were, on the average, one year longer than those given to otherwise similarly situated offenders who had no prior felony convictions, we can infer that judges have a policy of adding one year to an offender's prison term if he or she has a prior felony conviction. This is true, the argument goes, regardless of whether the judges have articulated or are even subjectively aware of such a policy.

The difficulty with this reasoning is that it may assume too much. It may be, although it seems unlikely, that individual judges do not have consistent practices from which implicit policies can be inferred. It is likely, however, that different judges have differing practices based on different reasoning. The key words are "on the average." It is possible that the average tendency of all judges is something with which every single judge would disagree, and which is without rational justification.

Suppose, for example, that a small minority of judges rely heavily on factors that pertain to the offender's ability to reform, while the majority of judges pay no attention to those factors because they believe that no offender should be incarcerated for the purpose of rehabilitation. When sentences imposed by all judges are aggregated, offenders whose characteristics indicate a high degree of capacity for rehabilitation may, on the average, receive prison terms that are slightly shorter than do offenders whose prospects of reform appear bleak. This slight difference in length of incarceration, however, may be correctionally insignificant. That is, it may make no difference in terms of the offender's progress toward rehabilitation. A judge who favors rehabilitation as a purpose of incarceration would have preferred a greater difference in the sentences, while another judge would have preferred no difference, but both might agree that the middle position is indefensible.

When policymakers fail to understand and control technology, technologists may become policymakers in the breach.

This is not to say that inductive modeling of the sentencing process cannot arrive at sensible policy positions. Our example is admittedly extreme and probably unusual. Nevertheless, it illustrates the point that an empirical model of the sentencing process may reveal central tendencies that do not make theoretical sense, quite apart from the possibility that it will reveal practices that are condemned by law, such as racial discrimination. It is a long step from the descriptive statements that constitute a model to the normative statements that constitute policy.

The researchers who developed the first sentencing guidelines invited the judges to adjust the guidelines models to suit their policy preferences. Although some judicial modifications were made, the judges' direct impact on the development of the models appears to have been limited. In an enterprise of this kind, the researchers inevitably make many decisions that have significant policy implications. Of course, the judges were free to reject any guidelines with which they disagreed, but it is doubtful that they were able to review effectively the many small but significant decisions that were made by the researchers. The problem lies in judges' limited ability to understand the statistical models. From our interviews with them, we learned that the judges did not consider themselves competent to evaluate the methods by which the guidelines were created. Under the circumstances, they were required to take on faith much that should have been reviewed and ultimately decided by them. When policymakers fail to understand and control technology, technologists may become policymakers in the breach.

This demonstrates a major difficulty in using empirical models of past practices to formulate public policy. Although perhaps the dimensions of the problems can be reduced by clearer, more careful explanations, the difficulty cannot be eliminated unless the policymakers have special training and experience. This may be taken as an argument in favor of special commissions whose members have or can acquire the expertise necessary for an intelligent evaluation of proposed models.

A detailed consideration of the process by which the Denver, Chicago, and Newark sentencing guidelines were created reveals that they do not have a sound empirical basis. Lack of documentation of modifications made during the implementation phase of the Denver guidelines project forces us to limit our conclusions regarding the Denver guidelines development process to the feasibility phase of that project. The guidelines researchers did not test the assumption, which is implicit in their regression analyses, that judges are homogeneous with respect to the weight they give various facts pertaining to the offense and the offender's background. Consequently, they may have used an inappropriate unit of analysis when they elected to aggregate cases decided by different judges instead

THE AUTHORS

William D. Rich is an assistant professor at the University of Akron School of Law. Previously he directed the National Center's study of sentencing guidelines, from which this article is taken. Rich was also a research associate for two other Center projects, one an evaluation of a jury utilization and management demonstration program, and the other a project investigating the link between learning disabilities and juvenile delinquency. Rich has worked extensively in the field of court research for Creighton University and the University of Denver College of Law. He has a J.D. degree from the University of Denver and is a candidate for an LL.M. at Harvard Law School.

L. Paul Sutton is a specialist in the study of criminal justice and is the author or coauthor of nine books on aspects of this subject. Currently an associate professor at the School of Public Administration and Urban Studies at San Diego State University and director of the program in criminal justice, Sutton was a member of the National Center's staff for several years, working primarily on the sentencing guidelines project outlined here and on a project investigating the relationship between learning disabilities and juvenile delinquency. Sutton's Ph.D. and M.A. degrees in criminal justice are from the State University of New York at Albany. He earned his B.A. in political science from the University of Kansas.

Todd R. Clear, an assistant professor in the School of Criminal Justice at Rutgers University, is also project director of the Seminars for Key Decision Makers, which is funded by the National Institute of Corrections. He was a member of the National Center's sentencing guidelines project staff on a part-time basis for several years. Clear is a specialist in classification and supervision practices in the fields of probation and parole and has published numerous articles and books in this area, most recently serving as coauthor of Community Supervision of Offenders. *He holds a Ph.D. in criminal justice from the State University of New York in Albany and an M.A. in criminal justice and a B.A. in sociology from Anderson College.*

Michael J. Saks is an associate professor of psychology at Boston College. Previously he was a research associate at the National Center for the sentencing guidelines project, a study investigating the relationship between learning disabilities and juvenile delinquency, and a project to implement the ABA's standards of judicial administration. He recently was a consultant to the National Center's project studying the use of scientific and technological evidence in litigation. Saks holds a B.A. in English and an M.S. in psychology from Pennsylvania State University and an M.A. and Ph.D. from Ohio State University.

of analyzing each judge's cases separately. Our analysis of the data we collected for the impact study suggests that different judges may in fact have different "implicit sentencing policies." Specifically, we found that judges are not homogeneous with respect to the regression of the decision whether to incarcerate on several offense- and offender-related variables, including some that were included in the guidelines. This implies that different judges consider different factors with different weights when they sentence. The pooling of cases across judges therefore cannot result in the discovery of *an* implicit sentencing policy, but rather a possibly meaningless average of several different implicit policies.

One solution to this problem is to construct individual, judge-specific models for presentation to those responsible for making sentencing policy. These should help to identify the variety of sentencing policies apparently followed by different judges, as well as to stimulate debate and perhaps eventually consensus. Unless the differences are resolved, a single set of empirically based guidelines would be either a fiction or an arbitrary choice.

In our opinion the procedure by which variables were selected for inclusion in the guidelines models were deficient on both conceptual and technical grounds. Variables were selected largely on *post hoc* empirical grounds rather than *a priori* theoretical grounds. Variables were screened initially on the basis of the magnitude and significance of their bivariate relationships with the dependent variable (i.e., sentence). This introduced the risk that some variables that are important to the sentencing decision were excluded because their true relationships to the sentences were disguised by the effects of other variables.

Those variables that survived the initial screening were subjected to stepwise multiple regression analysis. Stepwise multiple regression is a procedure by which variables are selected mechanically for inclusion in a model on the basis of their incremental predictive value. Stepwise regression is best suited to the situation in which we lack a comprehensive theory about the behavior being modeled and wish only to make the best (i.e., most accurate) predictions on the basis of empirical observation. Although it is true that there is no comprehensive theory of sentencing that would allow one to choose variables on purely theoretical grounds, when one develops sentencing guidelines one does not seek merely to predict judicial sentencing behavior by whatever means are available. The variables that best predict sentences are not necessarily the same variables that judges use or ought to use.

The failure to distinguish adequately between prediction and explanation is fatal to the justification of empirically based sentencing guidelines. If one wishes to describe the implicit policy of sentencing judges, mere predictive accuracy will not suffice; the model must embody a thorough understanding (i.e., explanation) of the judges' behavior. Stepwise multiple regression analysis of data collected from criminal case files is incapable of yielding an adequate explanation of why judges sentence as they do.

Technical criticisms of stepwise multiple regression also counsel against its use in specifying guideline models. The existence of intercorrelations (i.e., collinearity) among predictor variables affects their marginal contribution to the predictive accuracy of the model, which in turn affects the order in which variables are selected for inclusion in the model. Moreover, the primary criterion for the inclusion of variables (the R^2 increment) is sample-specific. That is, its magnitude depends partly on the variances of the predictor variables in a particular sample. Therefore, the variables' order of entry into the regression equation (or R^2 increment) is an inappropriate measure of the importance of variables.

Collinearity is a problem of insufficient information. Certain combinations of offense- and offender-related characteristics apparently do not occur with sufficient frequency in the real world for us to determine exactly which characteristics cause judges to sentence as they do. One solution, perhaps the only true solution, is to have judges impose sentences in hypothetical cases in which the variables have been systematically manipulated so as to avoid the problem of collinearity. This approach has its own difficulties, but it seems well worth exploration.

An examination of the regression results reported by the guidelines researchers reveals that their models had little predictive accuracy, quite apart from their validity as explanations of judicial sentencing behavior. Most models accounted for less than one-third of the variance (variation) in sentences. Thus, not only do the models fail to explain what causes judges to sentence as they do, and not only are the results of questionable generalizability, but the models do not even predict accurately the sentences in the samples on which they were developed. The point of this criticism, however, is not that the guidelines models yielded unusually poor predictions of sentencing decisions. Statistical models of judicial sentencing behavior, including those we developed to study the guidelines' impact, rarely explain much more of the variance in sentences than did the guidelines models. The point is that these kinds of statistical models, which provide only crude explanations of judicial sentencing behavior, do not constitute an adequate basis for the formulation of sentencing policy.

Other weaknesses in the sentencing guidelines research included (1) the omission of highly skewed variables (i.e., characteristics or events that occur infrequently but nevertheless may influence sentences in cases where they do occur); (2) the premature exclusion of legally irrelevant variables (e.g., race, sex, bail status) that may in fact influence sentence decisions; (3) the failure to account for conditional relationships between variables which might have been discovered by including interaction terms in the regression models; (4) the failure to consider nonlinear functional forms that may correspond more closely to the actual decision process than does the linear form; (5) the absence of important information from the case files and therefore from the data set; (6) inaccurate coding of data; (7) the use of biased samples; (8) the inclusion of nonconvictions such as deferred judgments in the sample of convictions despite the fact that, in practice, the prosecutor controls the decision to grant deferrals—perhaps on grounds that are different from those used by judges when they impose sentences; and (9) inadeqate measurement of variables.

Two aspects of the development of sentencing guidelines require special comment. First, the sample of cases used in the feasibility study to construct the Denver guidelines models was too small. The Denver regressions reportedly were performed on a sample of only 120 cases. Moreover, our reanalysis shows that listwise deletion of cases with missing observations would reduce the sample to less than 50 cases. With fourteen predictor variables, this would result in a cases-to-variables ratio of less than four-to-one. Any similarity between the results of those regressions and judges' actual sentencing practices would be little more than coincidental. The Denver guidelines' claim to a sound empirical basis cannot be sustained on these grounds.

The second aspect of the development of sentencing guidelines that requires special comment is the measurement of the sentence as a single variable. The guidelines researchers' findings imply, and our analysis confirms, that the determinants of the decision whether to incarcerate are different from determinants of length of incarceration. According to our analysis, the best predictors of the decision whether to incarcerate include the statutory class of the offense, the offender's most serious prior sentence, whether the offender was released on bail pending disposition of the instant offense, the offender's liberty status (i.e., on parole, probation, or bail at the time of the offense), and the offender's employment status. The best predictors of length of incarceration were the statutory class of the offense, whether a victim was injured during the commission of the crime, whether a charge reduction was granted, method of conviction (jury trial versus guilty plea or bench trial), and the number of counts of which the offender was convicted. In general, the decision whether to incarcerate seems to depend largely on the offender's criminal history and demonstrated ability to function in society, whereas the length of incarceration seems to depend on the seriousness of the offense and factors related to plea bargaining.

In view of these findings, the guidelines' use of a single set of variables (and a single set of weights) represents a significant departure from empirical reality. At least to that extent, the guidelines are not accurate representations of the way judges decide sentences. Thus, the decision to use a single set of variables to determine both the decision whether to incarcerate and the decision about length of incarceration cannot be justified on empirical grounds. It must be justified, if at all, on policy grounds. Although the use of a single set of variables is administratively convenient, the argument that the purposes of sentencing are well served by this solution remains unarticulated and probably will remain so until the focus of sentencing guidelines shifts from empirical science to the reasoned formulation of sentencing policy.

We conclude that the empirical approach by which the Denver, Chicago, and Newark guidelines were created provides only weak predictions and even weaker explanations of judicial sentencing behavior.

The role of empirical research in the formulation of judicial sentencing policy should be far more modest than that urged by the guidelines researchers.

Moreover, it is doubtful whether improved modeling techniques can capture the complexities and subtleties of sentencing in a way that would lend itself to translation into policy. Even if the technical problems could be overcome, the approach still would suffer from the problem of converting policy questions, which properly are decided by policymakers, into research questions, which necessarily are decided by researchers according to the needs of empirical science. The role of empirical research in the formulation of judicial sentencing policy should be far more modest than that urged by the guidelines researchers. A new approach to structuring judicial sentencing discretion is imperative.

The Effects of Sentencing Guidelines on Judicial Sentencing Discretion

The second part of our study was a quantitative assessment of the degree to which sentencing guidelines affected judges' exercise of sentencing discretion. We began by looking at the extent of judicial compliance with sentencing guidelines. In the absence of substantial compliance, there is no reason to expect that sentencing guidelines will accomplish their purpose. Next

we asked whether in fact sentencing guidelines have reduced disparity in sentences. Finally, we considered the possibility that sentencing guidelines caused an increase or a decrease in the overall severity of sentences, quite apart from their effect on disparity.

Compliance. The sentencing guidelines that were the subject of this study were deliberately designed to encompass 80 to 90 percent of the sentences that had been imposed prior to the guidelines' implementation. (This was accomplished by making the recommended ranges of sentences sufficiently broad that 80 to 90 percent of the cases in the guidelines construction samples were accommodated in the model.) Thus, even if judges ignored the guidelines, so long as they continued to sentence as before we would expect 80 to 90 percent of the cases to "comply" with the guidelines. The object of the guidelines was to reduce disparity by coaxing judges to sentence within the guidelines ranges, except in those cases that are so exceptional as to require a sentence that is outside the recommended range. The imposition of a noncompliant sentence was supposed to be justified by a written statement of the facts that made the case exceptional.

The reasoning of the guidelines researchers implies that, all other things being equal, if guidelines are successful we should observe a compliance rate that is higher than the 80 to 90 percent base rate, but not so high as to suggest a failure to distinguish truly exceptional cases where departures are warranted. Of course, there is no assurance that all other things are in fact equal. Nevertheless, low compliance rates suggest that the guidelines have had little if any effect on the exercise of judicial sentencing discretion.

In Denver, the average compliance rate on the decision whether to incarcerate was approximately 70 percent during the 24 months following implementation of the guidelines. This compliance level is decidedly lower than that anticipated by the guidelines researchers. Moreover, it must be remembered that the decision whether to incarcerate is dichotomous. Approximately 60 percent of convicted defendants in Denver are sentenced to incarceration. Therefore, the 70-percent compliance rate is not much higher than the (roughly) 60-percent "compliance" rate that would have resulted if the guidelines had called for incarceration in every case.

By retrospectively computing guideline sentences for preguidelines cases, we find that the proportion of cases in which the actual sentence (incarceration/nonincarceration) is the same as the guideline sentence is not significantly greater after guidelines than before guidelines. Thus, insofar as the decision whether to incarcerate is concerned, it appears that the Denver guidelines did not have an important influence on the exercise of judicial sentencing discretion.

The compliance rate for the decision whether to incarcerate was only slightly higher in Philadelphia, where judges complied in approximately 74 percent of the cases in the postguidelines sample. A comparison of this compliance rate with the "compliance" rate yielded by retrospective application of the guidelines to preguidelines cases reveals a small but statistically significant increase in the frequency with which actual sentences accorded with guideline sentences. This increase may reflect the guidelines' influence on judges' sentencing decisions, but there is an alternative explanation. Just before their implementation, the guidelines reportedly were adjusted to reflect a preexisting trend toward more severe sentences. Therefore, one would expect to observe greater agreement between guideline sentences and actual sentences after the adjustment (i.e., after implementation) than before, even if the guidelines had no effect on sentences.

Judges complied much less frequently with the guidelines' recommendations about length of incarceration than with the recommendations whether to incarcerate. In Denver, the minimum term of incarceration fell within the guidelines range in only 42 percent of the cases, while the compliance rate for maximum terms was 55 percent. In Philadelphia, the compliance rate for minimum length of incarceration was only 37 percent, while the compliance rate for the maximum term was 49 percent.

Total compliance, defined as compliance on the decision whether to incarcerate as well as on the minimum and maximum terms of imprisonment, occurred in only 47 percent of the Denver cases and 51.3 percent of the Philadelphia cases. These figures must be considered disappointingly low in light of the fact that the guidelines were designed to accommodate existing sentencing practices.

One of the most promising aspects of sentencing guidelines was the requirement that reasons be given in justification of departures from the guidelines. It was hoped that the articulation of reasons would provide grist for the development of a common law of sentencing. This hope was disappointed in Denver, where judges provided written reasons in only 12 percent of the cases in which they departed from the guidelines. The experience in Philadelphia was considerably different during the first six months of the guidelines' use: reasons were given in 64 percent of the cases in which the actual sentence differed from the guidelines recommendation. While most of the reasons given in Philadelphia were apparently legitimate, approximately 10 percent of them were so vague as to defeat the purpose of requiring reasons to be given; another 10 percent could be considered duplicative of guidelines criteria. In light of these difficulties, the reason-giving requirement cannot be considered an unmitigated success, but the approach seems to have potential if the requirement is taken seriously by the judges. The Philadelphia judges were able to articulate reasons that were not rote or mechanically applied.

The various measures employed in this study converge on a single conclusion: sentencing guidelines have had no detectable objectively manifested impact on the exercise of judicial sentencing discretion.

Sentence disparity. For purposes of this part of the study, sentence disparity is defined as differences in sentences attributable either (1) to the race or sex of the offender, or (2) to no identifiable characteristic of the offense or the offender. This latter category (2) includes differences between judges as well as inconsistency over time. Another kind of disparity, namely, differences in sentences attributable to the defendant's exercise of trial rights, is discussed at length in the final part of this study.

Using path analysis to isolate the direct effects of race and sex on sentences, we find that these effects were not significantly diminished after the implementation of guidelines. In other words, where racial or sexual disparities (i.e., differences between races or sexes that remained after other relevant variables were controlled) existed before guidelines, the guidelines do not appear to have reduced these disparities. On the other hand, neither does the inclusion of racially correlated factors (employment status and educational achievement) in the Philadelphia guidelines appear to have exacerbated racial differences. There is little evidence of *any* change in racial or sexual differences in sentences as a result of the introduction of sentencing guidelines.

The central goal of sentencing guidelines is to reduce between-judge differences in sentences as well as differences that result from individual judges' inconsistency over time by making sentences more consistently the product of the variables incorporated into the guidelines. The guidelines' efficacy in achieving this goal can be measured by regressing sentence, measured first as a dichotomy (incarceration/nonincarceration) and subsequently as a continuous variable (length of imprisonment in months), on the guidelines variables. This was done separately using the pre- and post-guidelines samples. We found that, in both Denver and Philadelphia, the pre- and post-guidelines samples were homogeneous with respect to the regression of sentence on the guidelines variables. Our interpretation of this result is that sentences were no more the product of the guidelines variables after guidelines than before guidelines. We conclude that sentencing guidelines failed to reduce sentence disparity, defined as statistically unexplained variation in sentences. The finding here is consistent with our previous finding of low compliance rates.

Finally, we considered the possibility that sentencing guidelines (inadvertently) caused an increase or decrease in the overall severity of sentence, apart from the issue of disparity. Plots of quarterly incarceration rates and average sentence lengths over time show that, in those instances where the pre- and post-guidelines measures of severity differed, the difference can be attributed to a trend that predated the implementation of sentencing guidelines. Therefore, especially in light of the previous findings, it seems highly unlikely that sentencing guidelines affected the overall severity of sentences in either Denver or Philadelphia.

The various measures employed in this part of the study converge on a single conclusion: sentencing guidelines have had no detectable, objectively manifested impact on the exercise of judicial sentencing discretion. Nonetheless, two caveats are in order. First, our data included only the first six months after the implementation of the Philadelphia guidelines. It remains possible that the Philadelphia guidelines will have delayed effects. Indeed, the guidelines may have their greatest impact as new judges are rotated onto the criminal bench.

The second caveat is that the lack of impact evident in Denver and Philadelphia does not imply that sentencing guidelines as such are inherently incapable of accomplishing their purposes. The guidelines evaluated in this report were adopted voluntarily by local judiciaries; consequently, they lacked the force and effect of law. Guidelines that are promulgated by an administrative body pursuant to legislative mandate, or by a state supreme court under its rule-making authority, and backed by legal remedies for noncompliance, might have significant effects on the exercise of judicial sentencing discretion.

Sentencing Guidelines and the Exercise of Discretion by Prosecution and Defense

The final part of our study was an inquiry into the relationship between sentencing guidelines and plea bargaining. The vast majority of convictions are obtained by means of plea bargains. Consummation of a plea agreement requires the consent of the defendant and often that of the prosecutor as well. The interdependence of the judge, the prosecution, and the defense poses serious problems for sentencing reform. One should not expect to alter the discretionary decisions of the judge without also affecting those of prosecution and defense. Sentencing guidelines are an attempt to structure, and thereby

alter, the sentencing discretion of the judge. To the extent that they are successful, they are likely to influence the actions of the prosecutor and the defendant. Conversely, the judges' dependence on the discretionary actions of the prosecutor and the defendant, if not adequately taken into account, may frustrate the achievement of the reform goals.

In Denver charge reductions are the principal subject of plea negotiations. Colorado's statutory sentencing structure gives the attorneys a great deal of control over the sentence through the mechanism of charge bargaining. Judges rarely participate in plea negotiations. The full-time assignment of prosecutors and public defenders to individual courtrooms promotes the formation of close-knit work groups. Consequently, the attorneys become familiar with "their" judge's sentencing practices. To the extent that sentencing discretion remains with the judge after a plea to a reduced charge has been tendered, the attorneys can anticipate the judge's decision.

The Denver judges made a deliberate effort to downplay the sentencing guidelines. Although judges received the completed guidelines worksheets stapled to the presentence report, no copies of the guidelines were given to the attorneys. With few exceptions, the attorneys considered the guidelines to be so unimportant that they did not avail themselves of the opportunity to examine them in the official court records. Their indifference can be attributed to (1) the guidelines' apparent lack of influence on the judges, (2) the attorneys' ability to predict judges' decisions, and (3) the fact that often the attorneys determine the sentence by negotiating charge reductions.

Thus, the Denver guidelines played no role in the plea bargaining process. In large measure, the established patterns of discretionary decision making precluded the guidelines from taking on a significant role in the Denver District Court.

Voluntary sentencing guidelines do not work; the increased confluence of interests among judges, prosecutors, and defense attorneys that might engender the support needed to make guidelines work is inconsistent with the ideal of adversary proceedings.

In Philadelphia, sentence bargaining predominates over charge bargaining, but recent prosecutorial restrictions on plea bargaining have forced the court to rely increasingly on an unusual and intriguing method of securing guilty pleas. Criminal cases are assigned to individual courtrooms partly on the basis of the defendant's willingness to waive a jury trial. Almost all jury trials are assigned to a handful of judges who are known for the extremely severe sentences they impose. By waiving a jury, a defendant can avoid being sentenced by any of the most severe judges. Our statistical analysis shows that defendants who are convicted by juries receive sentences that are, on the average, roughly ten years longer than those given to otherwise similarly situated offenders who waive juries. Not surprisingly, almost 97 percent of convictions are obtained without jury trials. The Philadelphia court capitalizes on the existence of sentence disparity by using it as a tool to dispose of cases.

The statistical analysis upon which the Philadelphia guidelines were constructed identified method of conviction as an important determinant of sentences, but it was omitted as a determinative factor in the guidelines because of doubts about its constitutionality. Had the guidelines succeeded in eliminating or sharply reducing sentence disparity, the court's unusual system of inducing jury waivers would have broken down. Instead, the guidelines were not adhered to by the judges, especially the severe judges who handle almost all of the jury trials. The guidelines' failure to accommodate the demands of the Philadelphia court's huge caseload preordained their ineffectiveness.

The Chicago sentencing guidelines never were fully implemented in the way their originators intended. Even in the early stages of implementation, judges regarded the guidelines as a study of sentence disparity rather than an attempted reform. They computed the guidelines worksheet, if at all, only *after* imposing sentences. The guidelines had no meaningful impact on the exercise of judicial sentencing discretion or on the plea bargaining process.

Nevertheless, an analysis of plea bargaining in Chicago yields some useful observations about the possibilities for sentencing reform. In February 1978, Illinois' new "determinate" sentencing law went into effect. In many instances the new law had the predicted effect of shifting sentencing discretion from the judge (by constricting the judge's choice of sentences) to the prosecutor through the mechanism of charge bargaining. This result, however, was avoided by the judges who took active roles in plea negotiations. By dominating the plea bargaining process and, where necessary, using bench trials to convict defendants of lesser included offenses in order to avoid legislative constraints on their discretion, some judges were able to maintain their positions in the courtroom workgroups and adhered to their established practices. If judges can overcome the constraints and discretion-shifting tendencies of a comprehensive legislative enactment, it seems unlikely that voluntary sentencing guidelines can achieve their goals without the active support of judges, prosecutors, and defense attorneys.

We conclude that in order for sentencing guidelines to structure sentencing discretion successfully and reduce disparities, at least two conditions must be met. First, sentencing guidelines must be given the force and effect of law. Voluntary sentencing guidelines do not work; the increased confluence of interests among judges, prosecutors and defense attorneys that might engender the support needed to make guidelines work is inconsistent with the ideal of adversary proceedings. Accordingly, the idea of voluntary sentencing guidelines should be abandoned, if indeed it was ever taken seriously.

Merely giving sentencing guidelines the force and effect of law will not, however, obviate the difficulties caused by the interdependence of judges, prosecutors, and defense attorneys. The need to induce guilty pleas will create pressure to charge bargain. As long as the prosecutor can negotiate reduced charges, the judge's role in sentencing will be diminished. Therefore, if sentencing discretion is to be structured without further distortion of the criminal process, method of conviction must be explicitly included as a factor determining sentences.

The inclusion of method of conviction as a determinative factor in sentencing guidelines will force a constitutional adjudication of the practice of sentencing differentially according to the method of conviction. The lame argument that this differential is a reward for pleading guilty rather than a punishment for exercising a constitutional right finally will be put to rest. It will become clear that economy is the only adequate explanation of differential sentencing practices.

Existing case law strongly suggests that the U.S. Supreme Court would uphold the inclusion of method of conviction as a determinative factor in sentencing guidelines. In any event, a constitutional decision will necessitate hard choices. Jury trials are expensive. If we are to abandon the inducement of guilty pleas and jury waivers by differential sentencing, society must be willing to provide the resources necessary to afford jury trials to all defendants who freely choose them. It simply will not do to condemn trial courts' differential sentencing practices while refusing to allocate them the funds necessary to provide more jury trials. The trial courts are in an untenable position, forced by financial circumstances to deny rights guaranteed by the law they are responsible for administering. To be sure, a decision to permit differential sentencing will not be without cost. We will be telling defendants in no uncertain terms that although the Constitution guarantees their trial rights, they may suffer additional punishment if they choose to exercise them. Such a decision is unlikely to engender the respect for law on which a free society depends, but the situation is hardly different now. Indeed, the present system is worse because it is also dishonest.

The inclusion of method of conviction as a determinative factor in sentencing guidelines will have salutary consequences in addition to the virtue of honesty. First, as a sort of truth-in-advertising act for criminal defendants, it would eliminate much of the illusory bargaining that often reduces the value of defendants' trial rights to nothing. Second, the size of the differential could be continually adjusted on the basis of empirical evidence so as to achieve the desired proportion of waivers by the least amount of coercion necessary. Finally, once charge reductions and other concessions from the prosecutor are no longer needed to induce guilty pleas, prosecutorial plea bargaining can be curtailed. □

END