

DEPARTMENT

METHODOLOGICAL PROBLEMS IN STUDIES OF SENTENCING

The law, especially Anglo-American law, gives as a rule wide latitude to the determination of the sentence that follows the conviction of a defendant.¹ The law may set the minimum or maximum sentence for the particular crime, it may set the range, or it may allow different types of sentences (fine, prison, etc.); at the very least, it provides a choice between two alternatives, for instance, after conviction of a capital crime, it now allows in most states a finding of death or life in prison.

As a rule it is the judge who determines the sentence, but in capital cases, and in some states in other cases, it is the jury.² In one state, California, it is the prison authority to whom the sentencing decision falls, the judge merely pronounces sentence in terms of the range set by the penal code. But wherever the discretionary power lies, discretion is almost unlimited. Very seldom will an appellate court impinge on this discretion and change the sentence.³

It is easy to see why the law hesitates to limit individual discretion; the factors that come into play here are so manifold and possibly so idiosyncratic that it is difficult to make general rules about sentencing. However, the great discrepancies that appeared among sentences for what seemed fairly comparable cases have had two effects. First, the

1. The situation is somewhat different on the European continent, where the penal code itself enumerates the mitigating and aggravating circumstances the judge must consider. See, however, note 11 *infra*.

2. In ten states, all in the South, the sentences in all criminal cases are determined by the jury within the range set by the penal code; in another three states the jury determines the sentence only for certain specified offenses. See H. KALVEN, H. ZEISEL, *THE AMERICAN JURY*, ch. 20 (1966).

3. See L. Hall, *Reduction of Criminal Sentences on Appeal*, 37 COLUM. L. REV. 521, 762 (1937); and *Penalty Penology of Appeal: Appellate Review of Legal But Excessive Sentences*, Note 15 VAND. L. REV. 67 (1962). Again, the situation on the European continent differs. There the sentence may be, and often is, appealed by either prosecutor or defendant, even when the finding of guilty is accepted.

courts became faced with the practical problem of explaining such discrepancies.⁴

The other effect of this wide discretionary power in sentencing was the stimulation of research designed to find out what in fact determined sentences. This research has been of two kinds: some studies aimed at finding out the whole range of determining causes, others were designed to test whether one particular cause, such as different regional standards, or different standards of individual judges made any difference.

THE RANGE OF CAUSES

The studies aimed at the general roster of causes have been of two kinds. Some related, by some sort of correlation analysis, the severity of a sentence to whatever was known about the case that might conceivably affect the sentence: the type of crime, the record and personality of the defendant, the personality of the judge, regional differences, and so forth. Such studies have to cope with the general difficulties of survey data analysis—retrospective analysis of a great many interrelated variables—and their findings are correspondingly tenuous.⁵

The other approach to learning about the range of causes is to go and ask the judges who impose the sentence. This method goes under the technical name of Reason Analysis.⁶ A study from the Office of the United States Attorney General is such a first attempt. It reports on the factors judges say tend to produce lenient and severe sentences.⁷ The exploration of what might be called negative sentencing, namely the discovery of the reasons for granting executive pardon from the death sentence, belongs in this group of studies. We have one good study of this type for the United States and one for the United Kingdom.⁸

4. Cellmates in prison, who view their respective crimes as being of comparable gravity, often find their sentences differ so widely as to raise serious questions of justice. To avoid such anomalies, the courts have recently begun to develop informal, internal guidelines for judges of the same jurisdiction, that implement the discretion within the sentencing range set by the law.

5. Examples of such studies are: G. W. Baab & W. R. Furgeson, Jr., *Texas Sentencing Practice: A Statistical Study*, 45 TEX. L. REV. 47 (1967); and E. GREEN, *JUDICIAL ATTITUDES IN SENTENCING* (1961).

6. See H. ZEISEL, *SAY IT WITH FIGURES* chs. 10 & 11 (5th rev. ed. 1968).

7. U. S. DEPARTMENT OF JUSTICE, *THE ATTORNEY GENERAL'S SURVEY OF RELEASE PROCEDURES* ch. 11 (1939).

8. *Executive Clemency in Capital Cases*, 39 N.Y.U.L. REV. 136 (1964); ROYAL COMMISSION ON CAPITAL PUNISHMENT, *MINUTES OF EVIDENCE* 1 (1949). See also KALVEN & ZEISEL, *supra* note 2, at ch. 36.

METHODOLOGICAL PROBLEMS IN STUDIES OF SENTENCING

To ask judges why they gave a certain sentence in a particular case is strangely enough an unnatural enterprise. Anglo-American law allows the judge, when he pronounces sentence, to be completely silent on why he fixed the sentence on a particular level.⁹

The most sensible way to begin such an exploration of reasons, is in a place where these reasons must already exist in articulated form, for instance, with the California Adult Prison Authority. That agency is charged with determining the sentences of all convicted prisoners within the wide frame provided by the law, and, presumably, must operate with some rules.¹⁰ Another source might be the opinions of the sentencing review courts of Connecticut and Massachusetts, the only states that have such an institution.¹¹

TESTING FOR ONE FACTOR

One will not be surprised to learn that the more successful approach to the sentencing problem has been in studies designed to test for the influence of one or the other *single* factor, that may or may not affect the level of sentencing. The simplest example is administrative statistics that compare the level of sentencing in different jurisdictions or regions.¹² Such crude statistics are meaningful to the extent that the severity of the crimes committed in these various jurisdictions is comparable. The best way to approach that problem is, therefore, to determine within each of these jurisdictions the level of severity of the committed crimes and base the comparison on whatever differences appear.¹³

9. The practice has been severely criticized by Professor Rupert Cross in his inaugural lecture at the University of Oxford, *Paradoxes in Prison Sentences*, 81 L.Q. REV. 205 (1965). Here again, the European practice is different: the penal code lists the circumstances that can alleviate or aggravate the sentence, and the court must say which ones pertained in the particular case. The Vera Institute of Justice, through what it calls the Bronx Sentencing Project, has made an interesting effort in this direction. It has analyzed actual sentences, given by fifty-seven judges of New York City's criminal court, and related them to the gravity of the offense and other factors that emerged from the presentencing report in each case. It thereby developed an empirical point system that served as a basis for sentence recommendation—and prediction.

10. See p. 623 above.

11. *Appellate Review of Primary Sentencing Decisions: A Connecticut Case Study*, 69 YALE L.J. 1453-78 (1960).

12. For instance: Maximum Sentences by State, National Prisoner Statistics, Prisoners Released from State and Federal Institutions in 1960, Table 55 (1963).

13. A fine example of this approach is R. HOOD, *SENTENCING IN MAGISTRATE'S COURTS, A STUDY IN VARIATIONS OF POLICY* (London, 1962).

NATURAL EXPERIMENTS

The perfect setting, of course, for such single-factor-studies is provided by the rare situations where there is a random assignment of cases to the various units under study, usually individual judges within the same court. Such situations provide the perfect conditions of a controlled experiment.¹⁴ Analytical difficulties can not arise, all one has to do is to compare the sentence pattern of Judge A with the sentence pattern of Judge B. Any observed differences can be safely assigned to the different personalities of the two judges.

Thus a study of the New York City Magistrate Court showed that, among the thirteen judges, one acquitted 73% of the defendants charged before him with public drunkenness, while one of his colleagues acquitted only 1% of his defendants.¹⁵

Random assignment of cases to judges also has the advantage of allowing a more detailed analysis, namely of whether differential sentencing applies also to any subcategory of the assigned cases, *e.g.* with respect to public drunkenness, to petty theft, etc. If a group of cases is assigned randomly, any included subgroup will also be assigned randomly.

RETROSPECTIVE SURVEY DATA

Studies of the effect of the judge's personality on sentencing are—with one exception, to be discussed later—the only ones that have the advantage of random assignment. Studies of other factors that do not have the advantage of the prospective design of a natural experiment had to fall back on retrospective survey data.

A fine example of this type of research design is a study undertaken by the National Defense Fund of the NAACP. That study was designed to find out whether Southern juries, in deliberating whether to impose the death sentence on a man convicted of forcible rape, permit the race of defendant and victim to affect their decision. Specifically, the

14. See H. ZEISEL, *supra* note 6, at ch. 7.

15. Everson, *The Human Element in Justice*, 10 J. CRIM. L. & C. 90 (1919). Later studies by F. J. Gaudet, *Individual Differences in the Sentencing Tendencies of Judges*, 32 ARCHIVES OF PSYCHOLOGY (1938) confirmed this idiosyncratic bias of judges. E. GREEN, *JUDICIAL ATTITUDES IN SENTENCING* (1961), based on data from the Criminal Court in Philadelphia, is the only study that claims to have found no such differences, but its statistical analysis is open to criticism.

METHODOLOGICAL PROBLEMS IN STUDIES OF SENTENCING

question was whether a Negro, convicted of having raped a white woman is, all other circumstances being equal, particularly likely to receive the death penalty.¹⁶ The Arkansas facet of the study established what is called a first order correlation between race constellation and frequency of the death penalty. In Arkansas, over a twenty-year period, Negroes, convicted of having raped a white woman, received the death penalty significantly more often than the defendants in other race constellations.

	<i>Negro Convicted of Raping a White Woman</i>	<i>All Other Convicted Rapists*</i>
Per cent of cases where defendant was sentenced to death	46%	14%

* This category consisted only of Negroes convicted of raping a Negro woman or white men convicted of raping a white woman. There were no cases of white men being convicted of raping a Negro woman.

On the basis of this information, the following theorem was appropriately introduced: it is possible that the first order correlation is misleading, since factors related to both the race constellation *and* the frequency of the death penalty, which could satisfactorily explain the original correlation, may exist. Suppose, for instance, more of the Negro-white rapists had used a weapon in their attack, and for that reason only showed a higher proportion of death sentences. Such an interrelationship would reveal the original correlation as spurious, that is, as not being caused by race prejudice, but by a legitimate factor. Therefore, the problem was to prove that the original correlation held not only for all cases combined, but also for every subgroup of comparable gravity, armed as well as unarmed.

The study then suggested a number of factors that could conceivably explain why the original correlation—while true for all cases as a

16. Forcible rape is still a capital crime in eleven Southern states, and the study covered those states. The results of the study have been presented for cases in Alabama, Arkansas, Florida, Georgia, Louisiana and South Carolina, and this paper focuses on the Arkansas part of the study. That facet is discussed and evaluated by the District Court at 257 F. Supp. 710 (E.D. Ark. 1966) and by the Court of Appeals at 398 F.2d 138 — U.S. —, *cert. set down for reargument*, October, 1969 Term (No. 622) (8th Cir. July 11, 1968) in the case of Maxwell v. Bishop. On December 16, 1968, the Supreme Court granted certiorari in that case, but limited the questions it would consider so as to exclude considerations of the issues directly related to the study (37 U.S.L.W. 3214). The study is more briefly described in *Matter of Sims and Abrams*, 389 F.2d 148 (5th Cir. 1967), involving the Georgia facet, and in *Moorer v. South Carolina*, 368 F.2d 458 (4th Cir. 1966), involving the South Carolina facet.

whole—could disappear if one compared the race-death penalty relationship for *comparable* cases of rape.

For each of the factors, the study determined first whether it was related to the relative frequency of the death penalty. In order to decide whether or not there was in fact a correlation, the p-values were computed, indicating the probability that the observed differences could have been obtained by chance. If that value reached the frequently accepted standard level of $p = .05$ (5% chance), it was decided that the two factors were correlated.

<i>Factor*</i>	<i>P - Value Correlation With Death Penalty</i>
Age of defendant95
Mental status of defendant95
Defendant has dependent children90
Defendant had record of criminal conviction10
Defendant had prison record01
Age of victim95
Victim has dependent children70
Contemporaneous other offenses05
Place of offense (indoor v. outdoor)95
Illegal entry to place of offense50
Display of weapon98
Seriousness of injury to victim95
Victim was stranger10
Defendant pleaded guilty30
Quality of counsel (appointed v. assigned)80

* Other factors were considered such as victim's reputation, but the record did not contain adequate data.

Only two factors, *prior prison record* and *contemporaneous other offenses* reached the required probability level. These two factors were then tested to determine whether they were also related to the race constellation.

METHODOLOGICAL PROBLEMS IN STUDIES OF SENTENCING

	<i>Contemporaneous Offenses</i>		<i>Prison Record</i>	
	Negro ag. White	All Others	Negro ag. White	All Others
	Percent receiving death penalty	67%	28%	73%
Number of cases	(9)	(7)	(11)	(15)

It turned out in each case that, although the number of defendants became small at this point, the original correlation prevailed.

The analysis was not carried further, because it was thought that the small number of available cases did not warrant it.

We turn now to another study of this kind, an offshoot of the by now famous bail experiment of the Vera Institute. Originally, it was designed to find out whether a relaxation of the bail rules, allowing a much greater proportion of defendants to be out of jail prior to their trial, would result in a higher nonappearance rate at the time of trial. As is well known by now, no such effect appeared, and the recommended procedure has made history. But this is another story, already told.¹⁷

Here we are concerned with a follow-up study, based on the data from the Vera experiment. Its aim was to find out whether a defendant who is free on bail at the time of his trial, is, all other circumstances being equal, treated less severely than the defendant who is still in jail at the time of his trial. It is one of those all too rare studies in which a costly set of primary data is put to a secondary use.¹⁸ Again, the study begins by establishing the crucial first order correlation:

	<i>Defendant Was at Time of Trial</i>	
	% on Bail	% in Jail
	Sentenced to prison	17
Lesser sentence or acquitted	83	39
	100	100

17. See H. ZEISEL, *The Law*, in THE USES OF SOCIOLOGY (P. Lazarsfeld ed. 1968).

18. A. Rankin, *The Effect of Pretrial Detention*, N.Y.U.L. REV. 941-55 (1965). Curiously enough, the analyst failed to perceive that an experiment designed to test the effect of variable *a* on variable *b*, is also an experiment testing the effect of *a* on any other factor.

Next, three factors were suggested that could possibly explain the first order correlation by revealing it as spurious, if they happened to be related to both the bail-jail situation (the alleged cause) and the prison—lesser sentence situation (the effect).

<i>Percent Detained in Jail</i>		<i>Percent Receiving Prison Sentence</i>
	<i>Criminal Record</i>	
61	yes	63
38	no	29
	<i>Amount of Bail</i>	
61	\$500 or less	51
35	over \$500	26
	<i>Quality of Counsel*</i>	
70	court assigned	51
17	private attorney	23

* Two more factors were considered, *family integration* and *employability*, but they were found to be so highly correlated with *quality of counsel*, that too few cases would be available in these refined categories. Hence *quality of counsel* stands in the analysis for all three variables.

All three factors emerge as potential explicators, since each turned out to be related to both factors. The analysis then proceeded by relating each simultaneously to the two factors of the original correlation. Here is the criminal record as an example:

BEING FREE ON BAIL AND SEVERITY OF SENTENCE BY CRIMINAL RECORD

	<i>No Record</i>		<i>Previous Record</i>	
	%	%	%	%
	on bail	in jail	on bail	in jail
Sentenced to prison	9	51	36	81
Lesser sentence or acquitted	91	49	64	19
	—	—	—	—
	100	100	100	100

METHODOLOGICAL PROBLEMS IN STUDIES OF SENTENCING

Consideration of the criminal record, it turns out, does not only not destroy the original correlation but strengthens it. The discrepancies between *bail* and *jail* are greater here than in the original correlation for the cases as a whole. Analogous tabulations of the other two factors yielded essentially the same results in that the original correlation was still maintained, although in slightly reduced form.

Not content with these results, the analyst went one step further, arguing that while none of these factors alone was able to destroy the original correlation, perhaps jointly they could. (This step was missing in the rape study discussed earlier.) The following table was then constructed:

BEING FREE ON BAIL AND SEVERITY OF SENTENCE WHEN NUMBER OF FAVORABLE CHARACTERISTICS IS HELD CONSTANT

	<i>Number of Favorable Characteristics*</i>							
	None		One		Two		Three	
	on bail	in jail	on bail	in jail	on bail	in jail	on bail	in jail
Percent Sentenced to Prison in Each Group	72	82	26	73	17	52	6	—**
Number of Cases	(18)	(107)	(68)	(110)	(122)	(62)	(67)	(2)

* A favorable characteristic is one that favors the defendant. *E.g.*, no criminal record, etc.

** Too few cases.

The table shows that on each of the four levels—each more favorable to the defendant than the preceding one—the defendants in jail are treated more harshly.¹⁹ In principle this type of table reconstructs the experimental design. Actually it suffers typically from attrition of cases in the marginal categories, while cases in the center groups bulge.

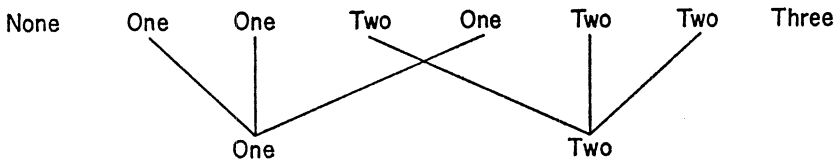
Secondly, this type of arrangement assumes that each factor has about the same effect, and that these effects are additive. The above

19. I should like to add a digression here. Miss Rankin's study shows that being in jail at the time of trial not only increases the sentence in case of conviction, but also increases the chances of being convicted. I have seldom seen a more disconcerting result of an empirical legal study; yet it caused not so much as a ripple in the law world. This is a melancholy comment on the powers of social research to change the real world.

table is, in fact, a contraction of the eight categories formed by the full combination of all three factors to four. We see now why the two

Complete Cross-Tabulation of "Favorable Characteristics"

Criminal Record				No Record			
\$500 Bail or More		Less Than \$500		\$500 Bail or More		Less Than \$500	
Assigned Counsel	Private Counsel	Assigned Counsel	Private Counsel	Assigned Counsel	Private Counsel	Assigned Counsel	Private Counsel



center categories ("one" and "two" favorable characteristics) bulge; each one represents in fact three categories.

Nevertheless, it is always advisable to see the *complete* cross-tabulation even if some of its cells turn out to be empty. First of all, techniques have now been developed for analyzing some of these incomplete tables;²⁰ secondly, it is only from such a table that one can decide on the proper eventual compromise solution.²¹

20. See, for instance, L. Goodman, *The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables With or Without Missing Entries*. J. AM. STATIST. ASS'N, Dec. 1968.

21. Finally, I should like to report here still another, hybrid approach to the problem. It was developed by Rupert Cross (*supra* note 9), and rests on asking judges what sentence they would give in cases submitted to them in a thumbnail sketch containing all the relevant features of the case. However, the point of the experiment, unknown to the judges, was the occasional addition of a potentially prejudicial circumstance, clearly irrelevant to the severity of the crime. Thus, one group of the judges would receive a case of negligent automobile homicide, while another group of

THE DEATH SENTENCE IN MURDER CASES

In the states that have retained the death penalty, it is, as a rule, the duty of the jury to decide whether the defendant is to receive it. And since the law here insists on giving the jury no guidance as to when it is to impose the death sentence, the quest for the jury's reasons is particularly difficult. Two major attempts have been made to explore this area. One proceeded by way of "reason analysis,"²² as part of the Jury Project of the University of Chicago Law School.²³ The other, more recently, by way of discriminant (multiple regression) analysis, of 238 first-degree murder cases in which California juries actually made the decision, gave the death sentence in 103 cases, and withheld it in 135.²⁴ Here the attempt was made to determine the extent to which any or all of 178 factors, for which data were made available on these 238 cases, moved the jury to give or withhold the death penalty. These factors, some attributes, some variables, covered five major areas: the defendant, the victim, the circumstances of the crime, characteristics of the trial, and the main actors in the trial—the judge, the prosecutor, and the defense lawyers. But the critical review of this important and complex enterprise must await another day.

—HANS ZEISEL

judges would receive the very same case with one innocent addition, such as that the guilty driver "a married man, had his mistress in the automobile." It is my understanding that the adulterers were "punished" more severely.

22. See note 6, *supra*.

23. See Kalvan and Zeisel, *THE AMERICAN JURY*, 1966, *A Somber Postscript: Decisions on the Death Penalty*.

24. *A Study of the California Penalty Jury in First-Degree Murder Cases*, 21 *STANFORD LAW R.* 1297 (1969).