POLICY EVALUATION AND RECIDIVISM

MARTIN A. LEVIN, *Brandeis University*

I. THE EVALUATION OF POLICY IMPACT AND SOME OF ITS GENERAL PROBLEMS

Non-experimental Analysis and Evaluation

During the past few years there has been a refreshing wave of studies of the relationship between policy inputs and outputs in American urban and state politics.[1] These studies have attempted to go beyond the analysis of the political processes of a unit of government, to analyze its relationship and that of other factors such as socio-economic characteristics to the policy outputs of that unit. They have gone beyond the analysis of "who governs?" to the analysis of "what difference does it make who governs?" and "what difference do certain socio-economic characteristics make?" (Wilson, 1964: 133). In other words, what are the consequences of these inputs for the life of the *average* citizen? These consequences have been analyzed

---

AUTHOR'S NOTE: *The research on which this study is based has been supported by a grant from the Law Enforcement Assistance Administration of the U.S. Department of Justice (Pilot Grant NI-70-065-Pa-19). Support for the final stages of its preparation was kindly provided by the James Gordon Foundation of Chicago. Earlier stages of the large project of which this is a part were supported by the Joint Center for Urban Studies of M.I.T. and Harvard and by a Research Training Fellowship of the Social Science Research Council. I wish to thank all these organizations for their support, but the analysis and views expressed here are my own. The fact that the National Institute of Law Enforcement and Criminal Justice furnished financial support to the activity described in this publicaton does not necessarily indicate the concurrence of the Institute in the statements or conclusions contained herein. I also wish to thank the following individuals for many helpful suggestions during the preliminary stages of this study: James Q. Wilson, Martin Shapiro, Eugene Bardach, Frank Levy, Sheen Kassouf, and Zeenas Sykes. Michael Furstenberg provided invaluable aid as my research assistant during the final stages of the study.*

in terms of the policy outputs and services of these governments in areas such as education, welfare, criminal justice, planning programs, and general social welfare measures.[2]

As the logical conclusion of these studies and the input-output framework, policy analysts ought to evaluate the *impact* or *outcomes* of these policy outputs and thus attempt to discover their ultimate consequences for society.[3] However, the nature of policy evaluation presents major problems: First, a particular policy output is often only one of a broad range of causal factors affecting the behavior at which the policy is directed. Second, the knowledge of the precise degree to which each of the causal factors affect this dependent variable typically is imperfect. For example, in evaluating the impact of alternative criminal court sentencing policies on recidivism rates, one is faced with these problems: Sentencing decisions are, at the most, only one of the many causal factors of recidivism. Also, it is difficult to ascertain the precise degree to which each of these causal factors affects recidivism rates. It is especially difficult to ascertain the precise degree to which criminal court sentencing decisions affect them.

The third problem in the nature of policy evaluation is that the range of causal factors is typically broad and very complex. Many of them are highly correlated. That is, real world policy analysis is often confronted with the problem of *multicollinearity*. This is the name given to the general problem which arises when some or all of the explanatory variables in a relation are highly correlated. It then becomes very difficult, if not impossible, to distinguish and assess their precise relative effects on the dependent variable.[4]

There have been two major responses to these problems in the evaluation of policy impact. The first has been evaluation by the use of regression analysis which is designed to predict effects in a multivariate situation. Controlled experimentation has been the second response.

In principle, regression analysis can hold several explanatory variables constant to ascertain the independent effect of another variable. In practice, however, it is only as effective as the nature of the data allows. If the explanatory variables are highly correlated (*i.e.*, multicollinear), then it is very difficult for regression analysis to assess the precise contribution of each of these variables.

The Coleman Report (1966) is a classic illustration of this problem in policy evaluation. It found that differences in family backgrounds of students account for much more variation in achievement than do school differences. However, this analysis has been criticized for greatly under-estimating the contribution of school quality.[5] In part this resulted from the difficulty in assessing the relative contributions of family background and school quality. This difficulty is a product of the high correlation between the explanatory variables: Good homes and good schools tend to occur together and weak homes and weak schools tend to occur together.

In short, regression analysis is only effective in assessing the precise contribution of several explanatory variables if the data are "internally controlled" (*i.e.*, if there is a good deal of independent variation among the explanatory variables). Significantly, multicollinearity also weakens inferences based on cross-tabulations.[6] Thus, multicollinearity puts analysis and evaluation "in the statistical position of not being able to make bricks without straw."[7]

The response of social scientists to the problem of multicollinearity generally is unsatisfactory, especially for the purposes of policy evaluation. They almost exclusively suggest obtaining additional data which hopefully will lessen the degree of correlation between the explanatory variables.[8] More information may be a suitable abstract solution to the problem, but it is often unsuitable for real-world policy evaluation.[9] Again, the Coleman Report (1966) is illustrative. It was based on an extremely large national sample (3,155 schools and 569,000 students), but this did not mitigate the correlation between the explanatory variables. There are simply very few cases of good homes and weak schools or weak homes and good schools, even in a national survey.

Therefore, policy evaluators often are faced with a "multicollinearity deadlock"[10] which in practice cannot be broken either by additional information or by the use of different statistical methods.[11] Moreover, multicollinearity seems to have special consequences for policy evaluation which have been overlooked by even the most careful analysts.[12] In policy evaluation, the program in question is often only one of several factors affecting the behavior at which it is directed, and often these factors are highly correlated. A program may have an effect, but because of multicollinearity it may not be possible to measure it precisely as distinguished from the effect of other intercorrelated factors. Thus, when evaluations of specific programs conclude that the programs make *no difference* (*e.g.,* crime and delinquency reduction programs, Operation Headstart, and other compensatory education programs),[13] this conclusion may be largely the result of the problem of multicollinearity.

For example, this occurred when the President's Commission on Law Enforcement and Criminal Justice (1967) analyzed the policy of improved street lighting. The proponents of this policy suggest that adequate — and particularly above adequate — street lighting will, first, deter certain types of street crimes by increasing the offender's risk of being detected and, second, enhance the probability of apprehending the offender. A study of Flint, Michigan's major improvement of its central business district lighting found that over a six-month period there was a 60% reduction in the number of all felonies and misdemeanors and an 80% reduction in larcenies. However, at the same time there was an increase in police surveillance of the area. Therefore, it is not possible to ascertain the precise effect of street lighting alone. In other words, two possible causal factors — improved lighting and increased police surveillance — were perfectly correlated, and it is impossible to assess the precise contribution of each. Both of these possible causal factors often are likely to occur together because they are the product of the same general force — the desire to reduce crime. Such a complex and collinear pattern is probably typical of policy programs and the factors surrounding them. Nevertheless, the findings of the Flint study and other studies of improved street lighting led the President's Commission to conclude that "there is no evidence that improved lighting would have a lasting or significant impact on crime rates," though they did add that "there is a strong suggestion that it might."[14]

In summary, the frequent absence of a satisfactory solution to the multicollinearity deadlock should lead one to be more cautious in rejecting a policy program as "making no difference." If such a finding seems to be indicated, one should then investigate the interrelationships among the independent variables to ascertain whether multicollinearity does in fact exist and, if so, to what degree. If it does exist, however, policy evaluators should not despair completely. First, as will be discussed in the following section, there are methods of policy evaluation other than regression analysis and cross-tabulation. Second, the experience of physicists is perhaps instructive. Heisenberg's uncertainty principle has caused difficulties in the field of subatomic theory, but on the whole physicists have made major theoretical strides despite this principle. (The uncertainty principle states that it is impossible to specify or determine simultaneously both the position and velocity of a particle with full accuracy. It is possible to fix either of these quantities as precisely as desired, but the more exactness in one, the increasing uncertainty in the other. This lack of precision results from the effect of the observation on the observed particle.) More importantly, the precision of applied science (*e.g.*, sending a man to the moon or pinpointing an ICBM target 3,000 miles away) has not been deterred significantly. The requirements of analysis seem to be different for pure science and applied science. Similarly, social scientists ought to be able to make strides in both pure and applied fields despite the complexity and frequent multicollinearity of the real world. Perhaps they ought to develop their own uncertainty principle: The closer one gets to the facts, the more difficult it is to offer confident generalizations. Moreover, they ought to become aware that the requirements of analysis are different for pure social science and applied or policy social science. The second part of this paper will attempt to indicate the fruitfulness of this distinction in a concrete case of policy evaluation and prescription.

Controlled Experimental Analysis and Evaluation

Controlled experimentation is an alternative response to some of the problems of evaluating policy impact. In randomly applying a program or treatment to a population, the various possible independent variables other than the program in question are controlled. The data are thus controlled by randomization at the outset rather than in an *ex post facto* manner, as in regression analysis. Also, in controlled experimentation the data can be manipulated so that they do not present problems for statistical analysis, such as multicollinearity. Finally, in controlled experimentation all variables which are present in that population are included, and it thus avoids the problem of failing to include them in the regression equation. This is especially important in policy evaluaton (*e.g.*, often we may not be fully aware of all the possible major causes of a social problem such as recidivism).

Recently policy analysts have had a few opportunities to conduct genuine controlled experimentation with social policies.[15] However, the difficulty with controlled experimental evaluations of policy impact seems to be less a problem of conducting them than a problem of the nature of their method-

https://doi.org/10.2307/3052911 Published online by Cambridge University Press

ology and interpretation: First, in controlled experimentation there is the danger of the "Hawthorne effect" occurring. The name comes from the intensive series of experiments conducted at the Western Electric Company's Hawthorne Works in Chicago in the 1920s to determine how various changes in working conditions would affect the performance of female workers.[16] The researchers found it was not significant whether the worker had more or less light but merely that she was the subject of attention and part of an experiment.[17]

A second, and until recently a less frequently analyzed set of problems in controlled experimentation, results from experimenter effects which produce a self-fulfilling prophecy and the special cases of experimenter bias and labeling. During the last ten years Robert Rosenthal and various associates have investigated experimenter effects and experimenter-subject interaction in various contexts.[18] One investigation focused on the effect of teacher expectations with experiments in which teachers were led to believe at the beginning of a school year that, on the basis of tests that had been administered toward the end of the preceding school year, certain of their pupils could be expected to show considerable academic improvement during the year. In actuality the children designated as potential "spurters" had been chosen at random and not on the basis of testing. Nonetheless, intelligence tests given after the experiment had been in progress for several months indicated that on the whole the randomly chosen children had improved more than the rest (Rosenthal and Jacobson, 1968b: 19-20).

Rosenthal and Jacobson had taken steps to make certain that the predictions about the children were not based on judgments derived from previously observed behavior. They thus explain this greater improvement as a function of a self-fulfilling prophecy which, in this case, was the teacher's positive expectations for these children. "The essence of the concept of the self-fulfilling prophecy," Rosenthal and Jacobson explain,

> is that one person's prediction of another person's behavior somehow comes to be realized. The prediction may, of course, be realized only in the perception of the predictor. It is also possible, however, that the predictor's expectation is communicated to the other person, perhaps in quite subtle and unintended ways, and so has an influence on his actual behavior.[19]

Rosenthal and Jacobson also tested the alternative explanation that these intelligence test results were a function of a Hawthorne effect rather than of a self-fulfilling prophecy. Perhaps the fact that researchers supported by federal funds were interested in this school led to a general improvement of morale and a greater effort on the part of the teachers. They are able to reject this alternative explanation because "a Hawthorne effect might account for the gains shown by the children in the control group, but it would not account for the greater gains made by the children in the experimental group" (Rosenthal and Jacobson, 1968b: 23).[20]

The lack of precision indicated by the Heisenberg uncertainty principle is another difficulty resulting from experimenter effects. This lack of precision specifically results from the effect of observation on the observed particle. In observing a system it is necessary to exchange energy and momentum with it. This exchange alters the original properties of the system.

The third of these methodological and interpretive problems in controlled experimentation is the initial selection of the population or universe from which the control and experimental groups will be randomly selected. (Of course, this is also a problem in non-experimental research.) The nature and characteristics of this population become a constraint on the generality of the conclusions drawn from the experiment. For example, if this population is not typical of the more general population at which the policy is to be directed or if it differs in even one or two major characteristics, the applicability of the experiment's conclusions for this more general population is clearly questionable. In policy evaluation there seems to be a tendency to select the population or universe on criteria of convenience and non-controversy. This often means those close at hand and those in political jurisdictions whose elected officials are willing to allow a policy experiment to occur. Such a population may not be typical of a larger urban area toward which the policy is generally directed.

II. THE EVALUATION OF ALTERNATIVE CRIMINAL COURT SENTENCING POLICIES

Non-experimental Analysis and Evaluation

This part of the study will evaluate the impact of alternative criminal court sentencing policies on reducing recidivism. Criminal court judges have a very high degree of discretion in sentencing decisions,[21] and thus there is often a good deal of variance among the decisions of several jurisdictions, even within the same state. For example, in 1966 in the state of California as a whole 32.0% of the convicted defendants in

Superior Court received probation. The range of frequency of probation among the state's thirteen largest counties was from 7.2 (Fresno) to 40.7 (Alameda., *i.e.*, the Oakland area). In Los Angeles County the frequency of probation was 37.0%, but in Orange County it was 12.5%; in San Francisco County it was 35.6%, but in Sacramento County it was 9.1% (Beattie and Bridges, 1970: 90). Similarly an earlier study that I conducted indicated during the mid-1960s approximately 49% of the convicted common felons received probation in Pittsburgh, while approximately 37% of them received it in Minneapolis. Moreover, this difference between these two cities is even greater when controls are introduced for factors such as race and prior record (Levin, 1970: Ch. 5).

Ultimately, this study is primarily concerned with the impact of criminal court policies on recidivism, but there seems to be no *a priori* reason to suspect that court policies would be the only factor, or even the predominant factor, shaping recidivism rates. The studies of recidivism that will be analyzed here are therefore those that deal with the impact of several variables, and the relative impact of each of these variables will be analyzed.

The studies of factors affecting recidivism all indicate that offenders who have received probation generally have significantly lower rates of recidivism than those who have been incarcerated.[22] They also indicate that of those incarcerated, the offenders who have received a shorter term of incarceration generally have a somewhat lower recidivism rate than those who receive longer terms. With a few exceptions, these differences persist when one controls for factors such as type of offense, type of community, the offender's age, race, and number of previous convictions. That is, the difference in recidivism rates for the two treatments generally remains the same for all types of offenders. However, for those with certain characteristics (*e.g.*, youthfulness, previous record) there are some significant variations in the overall recidivism rates when type of treatment is controlled (*e.g.*, for all those who receive probation the recidivism rates are highest for the youngest and for those with the greatest prior record).[23]

Beattie and Bridges' analysis in 1970 of recidivism rates of offenders who were either granted probation or were incarcerated by the Superior Courts of California's thirteen largest counties is the most comprehensive study to date of factors affecting recidivism (Beattie and Bridges, 1970).[24] It simultaneously analyzes recidivism for both those incarcerated and those granted probation, with controls for many factors other than type of treatment. It indicates that the "success" rate for those granted probation was 65.8% (2,148) after a one-year follow-up and 48.6% (2,561) for those sentenced to jail.[25] (The "success" rate cited here is Beattie and Bridges' "none" category which signifies no known arrest either for a new crime or for technical violation of probation or parole during the one-year follow-up period.) This difference between "success" rates for the probation and jail groups persists when the following factors are controlled: county, sex, age, race, prior record, offense; and when the following factors are controlled simultaneously: offense and age, offense and race, offense and prior record (Beattie and Bridges, 1970: 11-200).[26]

George Davis' earlier study indicates, after a four- to seven-year follow-up period, a "success" rate (no subsequent probation violations or arrests) of 67.1% (6,268) for all those granted probation and 56.7% (5,400) for those sentenced to jail for a short term ("probation plus jail") in fifty-six of California's fifty-eight counties (Davis, 1964: 12).[27] These overall rates for each type of treatment were not controlled for factors such as offense, age, race, and prior record nor were offenders incarcerated for longer terms included.[28]

Ralph England's study indicates, after a six- to eleven-year follow-up period, a "success" rate of 82.3% (490) for a sample of adult probationers sentenced in the Federal District Court of the Eastern District of Pennsylvania from 1939 to 1944 (England, 1957: 667-668). This study does not include any recidivism data on a comparable group of offenders who were incarcerated by this court. Also, England used a less stringent criterion of "success" than did the other studies, and the offenses in federal court are generally less serious than in a state court, which is the source of data in the other studies.[29]

In nine of eleven follow-up studies of recidivism rates of individuals placed on probation summarized by England, there was a "success" rate of 70% to 90% and in the other two it was between 60% and 70%. Again, the criterion of "success" used in most of them is less stringent than those in Beattie's or Davis' studies. However, aside from this, the validity of these findings is greatly bolstered by their uniformity and their breadth — they were carried out in five states and one European country over a thirty-year period (1921 to 1954).

Data from the California Department of Corrections (CDC) for individuals released after incarceration in California state prisons indicate, after a one-year follow-up period, a "failure" rate ranging from 24.7% to 34.2% with a median "failure" rate of 30.5% (9,226) for each year from 1958 to 1968.[30] (These data include no information on any type of "success" rate.) The criterion for "failure" used by the CDC is return to prison either with a new felony conviction or without one (*i.e.*, a technical violation). By contrast, when a similar criterion is applied to the Beattie and Bridges' data, the "failure" rate for those granted probation is only 10.9% (2,148).[31]

Since these two sets of data are both from California, they also enable us to examine possible differences in recidivism rates according to length of incarceration. All individuals in the Beattie and Bridges "jail" group were incarcerated for twelve months or less and their "failure" rate is 21.1% (2,561).[32] By contrast, all individuals in the CDC data were incarcerated for more than twelve months and, as noted, their median failure rate for these years was 30.5% (9,266). (The median term of incarceration for the CDC group ranged from twenty-four to thirty-six months during 1960 to 1968.)

A detailed 1970 study by Public Systems Incorporated (PSI) based on California Department of Corrections data for individuals released from state prisons in 1964 and 1966 indicates, after a three-year follow-up period, a "success" rate of 32.8% (1,423) and 33.6% (1,208) respectively, or about half that of the "success" rate of those California offenders granted probation in the Beattie and Bridges analysis (Kolodney, 1970: Vol. 2, III-7).[33] Also, a comparison of the PSI and the Beattie and Bridges data again indicates lower recidivism rates for shorter terms of incarceration: All of the Beattie-Bridges "jail" group had terms of twelve months or less and, as noted, their "success" rate was 48.6% (2,561); all in the PSI group had terms for more than twelve months, with the median term of incarceration of 30 months in 1964 and 36 months in 1961 and, as noted, their "success" rate for 1964 and 1966 was 32.8% (1,423) and 33.6% (1,208).

Charles Eichman's study of two groups of incarcerated offenders indicates a lower "failure" rate for those with shorter terms of incarceration. Eichman analyzed the post-release experience of a group who had been released from prison early by the state of Florida as required by the *Gideon* v. *Wainwright* "right to counsel" decision and a control group of full-term releases. (The two groups were carefully matched for similar characteristics such as prior convictions, type of offense, age, and occupational skill level. The small final sample was the result of rigorous selection among 406 prisoners for true matches. Upon release the *Gideon* early releases had been incarcerated for significantly less time than the full-term releases. Eichman found that after a twenty-eight month follow-up period the "failure" rate for the *Gideon* early release group was 13.6% (110) and 25.4% (110) for the full-term releases (Eichman, 1966: 48-56).[34] (Eichman's "failure" rate is based on subsequent incarceration and he found this difference in "failure" rates to be statistically significant.)

Daniel Glaser's monumental study of the federal prison and parole system indicates, after a four-year follow-up, a "success" rate of 52.2% (1,015) for individuals who had been incarcerated in the federal prison system. Glaser's sample was from federal prisons and thus includes offenses that are generally less serious than those in the other studies, which are based on state prison and probation populations. (See note 29, above.) Therefore, in comparison to them, the "success" rate of the Glaser study is probably somewhat of an overestimation (Glaser, 1964: 19-21).[35]

Glaser also describes three studies similar to his own which cover state prisons in California (1946 to 1949), Washington State (1957 to 1959), and Pennsylvania (1956 to 1958). They indicate that after follow-up periods of thirty-six months, six to thirty months, and approximately twenty-eight months, there were "success" rates of 28%, 49%, and 52% respectively (Glaser, 1964: 21-24).

Some of these studies analyzed the impact on recidivism factors other than the type of treatment prescribed by the court. Beattie and Bridges found that the younger the defendant, the more likely he was to repeat. For both those who received probation and those incarcerated, the youngest offenders had the lowest "success" rates and these rates increased for each age category (Beattie and Bridges, 1970: 14-15). They also found that Negro offenders have lower "success" rates than whites, for both offenders granted probation and those incarcerated (Beattie and Bridges, 1970: 15-28). The greater an offender's prior record, the more likely he is to repeat. For both those granted probation and those incarcerated, those with no prior record had the highest "success" rates, and these rates decreased for each level of a prior record (Beattie and Bridges, 1970: 16-29). They also found significant variation in the recidivism rates according to the type of offense. For both those granted probation and those incarcerated, those who had committed sex offenses and crimes against persons (homicide, robbery, and assaults) had the highest "success" rates respectively; those that had committed auto theft, burglary, and drug law violations had the lowest "success" rates respectively (Beattie and Bridges, 1970: 13, 24-25). The studies by George Davis, PSI, and Daniel Glaser have similar findings.

Thus, factors other than the type of treatment prescribed clearly have an impact on recidivism, but in almost all instances in the Beattie and Bridges data there is still a significant difference in recidivism rates for individuals with the same characteristics who receive different types of treatment. Indeed, with a few exceptions the *type of treatment* prescribed by the judge seems to have a greater impact than these characteristics or type of offense. However, one characteristic — the absence of a prior record — and two offenses — auto theft and drug law violations — seem to have a greater impact on recidivism than does the type of treatment prescribed by the judge: If an offender has no prior record, he will have a very high "success" rate no matter which type of treatment is prescribed by the court. For such offenders, Beattie and Bridges found that for those granted probation the "success" rate is 78.2% (687) and for those incarcerated it is 72.8% (377). Similarly, if an offender commits auto theft or a drug law violation he will have a low "success" rate no matter which type of treatment he receives. In a few other instances, the combination of a particular offense with another characteristic has a greater impact on recidivism than does the type of treatment.[36]

However, for all other offenses, including those that the data indicate have a major impact on recidivism, Beattie and Bridges' data indicate that the type of treatment has a greater impact than does the type of offense (*e.g.*, for burglary the "success" rates are low for both those granted probation and those incarcerated — 56.3% [304] and 43.8% [526] respectively — but it is significantly lower for those incarcerated.) Their data also indicate that for offenders with all other characteristics, including those that the data indicate have a major impact on recidivism, the type of treatment has a greater impact than do their characteristics, either individually or simultaneously.[37]

The regression analysis of the PSI study is the most sophisticated and careful effort thus far to assess the relative impact on recidivism of type of treatment, type of offense, and offender characteristics. However, the PSI study only analyzed an incarcerated population, and even within this limited population, some of the analysis is plagued by multicollinearity. It concluded that "at the 90% level of confidence, the variables which are associated with the response (*i.e.*, no recidivism) are, in order of their contribution, prior record, class, narcotics history, ethnic, (*i.e.*, racial) group, base expectancy, and age. Prior record, class and narcotic history are by far the most important variables. . . . The variable of primary interest, time served (in incarceration), 'fell out' of the model. This variable has no effect or is not associated with the probability that an individual is clean (*i.e.*, no recidivism)" (Kolodney, *et al.*, 1970: III-27).

It should be emphasized that this conclusion only applies to one type of treatment — incarceration — but there are reasons to be hesitant in accepting it even with respect to incarceration. First, the cross-tabulation analysis presented in the PSI study itself indicates that when type of offense is controlled, for most offenses there are significant differences in the recidivism rates for those incarcerated for "short" or "long" terms (Kolodney, *et al.*, 1970: III-18). Admittedly, of course, cross-tabulation is less powerful and less revealing than regression.)

Second, there are reasons to suspect that there was insufficient variation among the data points for the independent variable of "time served in incarceration" to properly assess its potential contribution to recidivism rates.[38] The PSI study uses the labels "short" and "long" terms of incarceration, but in fact, there are too few genuinely short terms of incarceration (*e.g.*, twelve months or less or even eighteen months or less) to test whether a short term has any impact on recidivism. Evidence for the latter possibility comes from the comparison noted above of the "jail" group data in the Beattie and Bridges study (those incarcerated twelve months or less) and the PSI sample (all of whom were incarcerated for more than twelve months and for whom the median term was thirty months in 1964 and thirty-six in 1966) which indicated lower recidivism rates for the "jail" group.

Third, the PSI study is plagued by multicollinearity. Several of the independent variables to be tested — such as prior record, length of incarceration, and ethnic group (*i.e.*, race) — appear to be highly intercorrelated.[39] This makes it difficult to assess their relative impact on recidivism with true precision.

Significantly, several of the independent variables in the Beattie and Bridges analysis also appear to be highly intercorrelated. This may weaken some of the conclusions based on their data concerning the relative impact of variables other than the type of treatment prescribed by the judge.

In summary, the non-experimental analyses of the factors shaping recidivism seem to indicate that on the whole the type of treatment has a major impact. However, they also indicate that other factors, such as type of offense, prior record, race, age, and narcotics history, also have a major impact. These analyses also indicate that on the whole those offenders who are granted probation generally have significantly lower rates of recidivism than those who have been incarcerated. This pattern generally tends to persist when offender characteristics and type of offense are controlled.

However, this general finding of lower recidivism rates for those granted probation, even when these other factors are controlled, does not necessarily indicate that the lower rates are a specific function of this type of treatment. Instead, this relationship may be largely an artifact of the court's decision-making process. It is possible that those granted probation have lower recidivism rates because, first, those individuals with "favorable" offenses and characteristics (*e.g.*, the absence of a prior record) are generally granted probation and, second, those individuals with these "favorable" offenses and characteristics are most likely to have lower recidivism rates.

In short, the judge's decision concerning type of treatment to prescribe tends to coincide — that is his intention — with the actual correlation between offender characteristics and recidivism. Indeed, the judge usually bases his decision on offender characteristics and type of offense. Thus it is possible that it is the offender's characteristics rather than anything inherent in the type of treatment, or anything inherent in being given one's freedom when probation is granted, that is the primary influence on recidivism. This suggestion would apply in an analogous manner to those incarcerated whose higher recidivism rates may be largely a function of their "unfavorable" characteristics, such as a serious prior record.

The data analyzed above, which tentatively indicate that a few characteristics and types of offenses may have greater impact on recidivism than the type of treatment received, in part tend to support this suggestion. On the whole, however, it does not seem possible to test this suggestion properly because of insufficient variation among the data points for several independent variables. For example, there are very few individuals with no prior record who are incarcerated; or, conversely, there are few Negroes with the following combination of characteristics who are granted probation: a serious prior record and the commission of a drug law violation. Moreover, for the purpose of policy evaluation and prescription, the possibility that the relationship between type of treatment and recidivism rates may be an artifact of the court's decision-making process may not be fully relevant. For this purpose it is insufficient to simply ascertain which factor is the "best predictor."

Policy makers need information about the explanatory factors over which they have some control. These factors may predict an outcome less perfectly, but they will probably give the policy maker greater ability to affect the outcome. A judge, or any other policy maker, can do little to change an offender's age or his number of prior convictions, but he can prescribe the precise type of treatment (probation or incarceration) which he will receive. The factors influencing educational achievement which are analyzed in the Coleman Report are another example of this pattern. Even if a student's family background is the best predictor of educational achievement, it is difficult for policy makers to influence this factor. By contrast, they do have some control over the size of his class in school, which in Coleman's analysis seems to have been a less important predictor of educational achievement. This pattern again seems to indicate that the requirements of analysis are different for pure social science and applied or policy social science.

The problems caused by the possibility that an apparent relationship is an artifact of the treatment process being analyzed are endemic to the analysis of non-experimental data. For example, if there is this insufficient variation among the data points of some of the independent variables, any type of statistical controls are of little help. An alternative method of analyzing the relationship between type of treatment and recidivism is conducting a controlled experiment. In this way the decision to grant probation or incarceration is not "contaminated" by a real decision maker.

Experimental Evaluations of Alternative Sentencing Polices

A controlled random experiment can isolate the effect on recidivism of the alternative types of treatment as opposed to the effect of a type of treatment linked to a type of individual — one who has been directed to that type of treatment by a judge. The various possible independent variables other than the program in question are controlled through a random application of that program or treatment to a population. The data are thus controlled by randomization at the outset rather than in an *ex post facto* manner (*e.g.*, regression analysis) as in non-experimental research.

The California Youth Authority has recently been conducting a controlled experiment in the cities of Stockton and Sacramento to evaluate the effectiveness of alternative treatment programs for convicted juveniles. At the level of general strategies for policy evaluation, the results and methods of this experiment — "The Community Treatment Project" (CTP) — seem to be indicative of both the potentialities and some of the drawbacks of experimental methods of evaluation (Warren, 1967). At the level of specific evaluation of alternative sentencing policies and specific policy strategies for reducing recidivism, the CTP results and methods are very useful and suggestive.

The CTP experiment involves an initial screening of convicted juvenile delinquents. The remainder are then randomly assigned either to an experimental group which is returned to the community (*i.e.*, receive probation) and receives intensive counseling, or to a control group which is assigned to California's regular juvenile penal institutions.[40] After a follow-up period of fifteen months the "failure" rate for the experimental group was 28% (134) and 52% (168) for the control group; after twenty-four months the respective "failure" rates were 38% and 61%. ("Failure" was defined here more inclusively than in the studies described above, such as Beattie-Bridges. It consisted of parole revocation which included "serious" violations [*e.g.*, new felony convictions and/or new incarceration] *and* "technical" violations which did not always involve an arrest. This may explain the lower "failure" rate — 10.9% — for the probation group in Beattie-Bridges, which becomes 34.2% when the Beattie-Bridges data are analyzed according to the CTP definition of "failure.") Personal and attitudinal change as reflected in psychological test scores were also measured during this period. The experimental group was also more "successful" according to this standard.[41]

However, as is often the case in experimental evaluation, the CTP experiment seems to have been flawed in four significant respects. First, the initial screening eliminated about 25% of the convicted male juveniles (and 10% of the females) for whom institutionalization was deemed mandatory because they were involved in serious assault cases or because there was community objection. This clearly limits the generality of the conclusions that can be drawn: These data indicate that recidivism is less likely if offenders receive probation, but we do not know if this applies to the most serious offenders.

Second, there seems to have been ambiguous specification of the independent variables in the creation of the experimental design. Those in the experimental group receive both probation *and* intensive counseling. Thus there is no way to ascertain which of these aspects of their treatment is related to their lower failure rates. To do this an additional experimental group should have been created which received probation but no counseling at all.

Third, because of the nature of the supervision of the experimental group it seems very possible that its lower "failure" rate is to some degree a function of experimenter effects such as the Hawthorne effect and the effects of a positive self-fulfilling prophecy and positive labeling. (An incidental element, but not the essence, of the second flaw is that one of the ambiguously specified variables — counseling — is in itself somewhat of a Hawthorne effect.) The youths in the experimental group receive *intensive* attention from a "community agent," *i.e.*, a probation officer) whose entire caseload is twelve youths, compared to a normal caseload of from four to eight times that number. The youths see the agent from two to five times weekly, either individually, in a group, or in family meetings. They receive *special types* of attention which usually have a *group-oriented focus*: group and family therapy sessions, various group activities, and school tutoring. Much of this activity focuses around a program center which resembles a settlement house.

The youths are not only aware that they are receiving intensive, special, and group-oriented attention and that they are part of an experiment, but it seems possible that they are also aware that they are "supposed to act better" because they have had this "extra break of not being incarcerated."

https://doi.org/10.2307/3052911 Published online by Cambridge University Press

This would seem to create a positive labeling effect[42] which could lower the "failure" rate for this group. This would be the converse of the often stated, though rarely systematically proven, view that incarceration and all the official and unofficial stigma attached, creates a negative labeling process which increases the "failure" rate of those labeled "prisoner" and "ex-con." Also, the decision to revoke probation for the experimental group is made by the community agents themselves.

This is not to suggest that the agents fail to attempt to uphold the standards of scientific objectivity that are necessary in an experiment. The evidence concerning their intentions is clearly to the contrary, but the issue here is the possibility of a more subtle and unconscious factor such as the agents' expectations and their effects. It seems quite likely that the agents expect the experimental group to do better and convey this expectation to the youths. (For example, the reports of the CTP experiment note that the agents are all probation officers who generally believe in probation, especially if it can occur in "ideal" and intensively supervised circumstances such as those in the CTP case). Thus the agents possibly contribute to a self-fulfilling prophecy. (Also, the CTP reports clearly state that the agents often do not make probation revocations for minor misbehavior. They do, however, often *"suspend"* the probation, which generally only involves serious warning. The reports state that in practice once the suspension is made, full revocation rarely follows. Although there is no direct evidence, the agents may be unfairly and unscientifically lenient to the experimental youths in these situations. Even if they are not, their behavior may still affect the "failure" rate through their expectation that the youths will "come around" and avoid revocation because of this new "second chance" they have had.)

The fourth flaw in the CTP experiment seems to have been the initial selection of the population or universe from which the experimental and control groups were selected (after the screening of the assault cases). The nature and characteristics of this population — convicted first offenders *from Sacramento and Stockton* — seem to weaken significantly the generality of the experiments. Neither city is typical of the large and heterogeneous urban areas from which the largest proportion of offenders come. Both cities are relatively small in comparison to Los Angeles, San Francisco, and Oakland; they are not heavily industrialized, and they do not have large Negro populations (though Sacramento has modest numbers of Mexican-Americans and Stockton has a sizeable number).[43] Indeed, the important and easily obtainable variables of race and ethnicity are not mentioned in the CTP experiment.[44]

The primary goal of both the CTP experiment and this paper is policy evaluation leading to prescription. As I have tried to indicate, the requirements of analysis seem to be different for pure social science and applied or policy social science. Thus, two of the scientifically flawed aspects of the CTP experiment are nevertheless quite promising possibilities for the goal of policy evaluation and prescription. However, let me emphasize that I clearly do not mean that invalid methods or findings should be tolerated when the investigator is primarily interested in policy evaluation and prescription. Accurate analysis and evaluation is the essential foundation of policy analysis. Yet as I will indicate, findings that are the product of somewhat less than perfectly controlled analysis may be of great heuristic value to the policy analyst. (The policy analyst's boldness and tolerance for uncertainty and imperfect findings ought to be tempered, however, by the awareness that his responsibility is even greater than a pure scientist's. The policy analyst's errors are much more costly — especially in immediate terms — than those of the pure social scientist. If a researcher is in error concerning the degree of pluralism in city X, then our understanding of the city's political process is faulty. However, if a rearcher is in error concerning the impact of program X on a population and his evaluation is acted upon, then many resources will be misallocated and it is possible that the population may be deprived of a potentially beneficial program.)

First, perhaps, the initial screening out of about 25% of the convicted male juveniles (and about 10% of the females) for whom institutionalization was deemed mandatory because of their assault background, points to a general policy prescription: Screen out such cases, and then grant probation to all other juvenile first offenders. According to the CTP findings, probation leads to less recidivism. Thus probation for all but those screened out by the above criteria should significantly lower the present "failure" rate.

Second, perhaps the possibility that the CTP's low "failure" rate for the experimental group is to some degree the function of experimenter effects can be utilized as an explicit and intentional positive policy. Indeed, though it is rather bold, a possible policy prescription flowing from this flawed aspect is: After an initial screening out of assault cases, all juvenile first offenders should be granted probation and assigned to an explicitly and intentionally "Hawthorne" and "positive self-fulfilling prophecy" program. The community agents would intentionally have expectations of "success" for these youths, who also would be positively labeled (as being an "experimental" participant).

The CTP findings indicate that this could significantly lower the present overall "failure" rate (*i.e.*, the combined rate for both offenders who are granted probation without a special program and those who are incarcerated). Investigations in the literature on experimenter effects ranging from Rosenthal's in the classroom and the animal laboratory to the use of "placebo effects" in medical science indicate that positive expectations, prophecies, biases, and labeling can be conveyed to a subject and can affect his behavior positively. Admittedly, however, institutionalizing the feeling of being in an experiment and being the focus of special attention is more difficult, especially for large numbers, but it is clearly possible. Precedents for the positive institutionalization of experimenter effects exist. The original Hawthorne experiments were directed toward this end, and in fact they greatly changed policies for employee-management relations. Similarly, Rosenthal and Jacobson's investigations have been directed toward the creation of programs for teacher training and classroom strategies that institutionalize positive expectations.

The policy prescription suggested here also must be considered in terms of the realistic policy constraint of *alternative costs*. At the present cost data are available only for Phase 2 of CTP experiment (see note 44 above), but they can give us an approximate idea of the costs of CTP Phase 1 and of the alternative costs of the policy suggested here. In Phase 2 the probation officers have caseloads of fifteen youths per officer (the caseload is twelve in Phase 1), and this costs $150 per month per boy which is three to four times as much as regular probation. However, it is still less than half the average monthly cost of incarcerating an offender. Phase 2 handles a group that is larger than the capacity of one of the new institutions that the Youth Authority is building at a cost of six to eight million dollars. Those who have criticized Phase 1 of CTP as being financially impractical for wide application have not made this type of comparison. (This type of comparison does not consider the probabilities of the cost to society if the offender recidivates while on probation. This and its alternative will be discussed below.)

Some Policy Implications

Both the non-experimental and experimental data analyzed above seem to indicate that, on the whole, those convicted individuals who are granted probation have lower recidivism rates than those who have been incarcerated. However, offender characteristics and type of offense committed — especially certain characteristics and certain offenses — also seem to have a significant impact on recidivism. Also, the experimental data seem to suggest that if probation programs can intentionally and explicitly develop "Hawthorne" effects and effects of positive self-fulfilling prophecies and positive labeling, then it may be possible that recidivism rates can be kept relatively low. For example, if there is initital screening of offenders to eliminate the most serious and dangerous offenders, this program of an intentional "Hawthorne effect" may be able to keep the "failure" rate below 30%. (In a later analysis there will be an attempt to estimate statistically the maximum threshold at which probation can be granted without significantly increasing the "failure" rate. What is the threshold for "good probation risks" and what are their characteristics; how many individuals would have to be initially screened out and what are their characteristics?)[45]

To convert these findings into policy guidance for criminal court judges and to apply them to the evaluation of a specific set of courts, the question of the goals of the criminal court must be analyzed. In addition to reduced recidivism, these seem to include maintaining order and stability in society, maintaining the freedom of the individual, satisfying a common notion of justice (*i.e.*, equality and consistency of treatment), maintaining an image of the court as a fair institution, maintaining the "declarative" nature of the criminal law (*i.e.*, the criminal law is in large part more intended to be a list of acts that society wishes to "declare" inappropriate rather than a list of acts against which it wishes full enforcement), and maintaining a favorable cost-effectiveness outcome for the courts' decisions. There is obviously considerable tension among these goals, especially among several of them and the goal of reduced recidivism. The following brief examples will illustrate this tension.

First, lower recidivism rates may be associated with a policy of probation such as the one proposed in the critique of the CTP experiment. This policy probably would satisfy the goal of reduced recidivism more than would increased incarceration. Nevertheless, it also risks significant short-run sacrifices in the goal of order and stability in society because it gives freedom to many convicted individuals who have a reasonably high probability of recidivating. Incarceration may have a small or negative effect on reducing recidivism. However, by denying the freedom of some individuals — especially those with a reasonably high likelihood of recidivating — it does tend to satisfy the goal of maintaining order and stability in society, at least in the short run. There is almost a zero probability of an offender recidivating while incarcerated. (The policy of probation suggested in the critique of CTP would mean a low number of incarcerations, and thus it probably would also involve sacrifices in the achievement of the goal of maintaining the "declarative" nature of the criminal law.)

Second, a policy to reduce recidivism may involve sacrifices in other goals even if it does not involve granting probation more frequently. For example, from the findings of this paper one could derive the following policy to reduce recidivism: Incarcerate, until they reach the age of 30 or 35, all individuals who commit their second felony offense. It is likely that this would reduce recidivism because the data indicate that after this age there is a sharp reduction in the probability of recidivating. However, this policy probably would contribute to the image of the court as an unfair institution. It would also involve sacrifices in the goal of maintaining a favorable cost-effectiveness outcome for the courts' decisions because of the immense capital and maintenance costs of incarceration. Indeed, the same amount of reduced recidivism achieved by this policy of incarceration until the age of 30 or 35 could probably be achieved by a probation policy at almost one-half the cost. (In a later analysis there will be a detailed examination of a cost-effectiveness comparison of probation and incarceration.)

Third, a brief evaluation of the sentencing decisions of the criminal court judges of Pittsburgh and Minneapolis indicates the difficulty in evaluating the most effective policy to reduce recidivism. An earlier study that I conducted indicated that sentencing decisions are more lenient in Pittsburgh than in Minneapolis. White and Negro defendants receive both a greater percentage of probation and a shorter length of incarceration in Pittsburgh. This pattern persists when the defendants' previous record, plea, and age are also controlled. Although both white and Negro defendants receive more lenient sentences (*i.e.*, more frequent grants of probation) in Pittsburgh, in both cities whites receive more lenient sentences than Negroes. However, this difference in the direction of greater leniency for whites is very small in Pittsburgh, while it is large in Minneapolis. Also, in Minneapolis defendants with a prior record receive a much lower percentage of probation and a much longer length of incarceration than do defendants with no prior record. In Pittsburgh, on the other hand, defendants with a prior record (with the exception of Negroes in a few categories) generally receive only a slightly lower percentage of probation and only a slightly longer length of incarceration than defendants with no prior record.[46]

On this basis one might conclude that the Pittsburgh judges' decisions, on the whole, tend to contribute more effectively to reduced recidivism because they grant probation more frequently. However, their frequent grants of probation for individuals with a high probability of recidivating (*e.g.*, those with a prior record and Negroes) probably does not effectively contribute to reduced recidivism. By contrast, the Minneapolis judges' generally severe decisions for these specific individuals may contribute to reduced recidivism more effectively. (Elsewhere I have attempted to evaluate the decision making of these two courts in terms of the multiple goals of the criminal court [Levin, forthcoming: Ch. 10].)

In summary, there is considerable tension among the goals of the criminal courts, as usually is the case with basic institutional goals and values. Indeed, few important goals and values in society can be simultaneously maximized. It is this tension which makes a consideration of these goals and values so fascinating and perplexing. However, in terms of the single goal of reduced recidivism, this study has attempted to offer more empirical guidance to decision makers and policy evaluators. Yet to achieve this goal, policy makers must also look beyond the criminal courts. As this study has indicated, factors other than court decisions also have a major impact on recidivism. The courts cannot and probably should not affect these factors.

NOTES

[1] See Wilson (1968) which includes or refers to many of these studies. See also Jacob and Vines (1965), Dye (1966), Jacob and Lipsky (1968), and Levin (1970); and Fry and Winters (1970) and the various studies referred to therein. For analysis of the impact of Supreme Court policy, see Muir (1967), Wasby (1970) and Becker (1969).

[2] For a more complete description of this approach and its theoretical underpinnings, such as David Easton's *The Political System*, see Levin (forthcoming: Ch. 1).

[3] For example, in the area of input-output analysis of comparative politics, Pennock (1966) argues that an approach that focuses on "outcomes" (the ultimate consequences for society of policy outputs) "deserves a certain priority" because "the test of anything in terms of what produces seems to make sense."

[4] For a technical discussion of multicollinearity, see Johnston (1963: 201-207).

[5] See Bowles and Levin (1968), which also deals with other methodological and statistical problems beyond the scope of our discussion here.

[6] See Tufte (1969: 653) and Blalock (1963) for examples and discussion of multicollinearity in cross-tabulation analysis.

[7] See Johnston (1963: 207). It must be noted, however, that multicollinearity is a statistical rather than a mathematical condition. Thus, one should think in terms of the problem's *severity* rather than its *existence* or non-existence. Also, despite its frequent presence, especially in policy data, multicollinearity is neither always severe nor always present. Farrar and Glaubner (1967: 94) suggest that the problem becomes severe when the explanatory variables are not just correlated but are also highly correlated (*e.g.,* greater than .75), when it is difficult or impossible to obtain the additional information to mitigate this high intercorrelation, and when in addition to these two conditions there are less than twenty data points.

[8] This is the standard response of econometricians (Johnston, 1963: 207; Farrar and Glaubner, 1967: 92, 106). This strategy has led them to frequently use cross-sectional over time-series data which has a high informational content, and they have apparently been successful at times in surmounting the problem of multicollinearity. See for example Prais and Houthakker (1955), Meyer and Kuh (1957), Orcutt (1961), and Stone (1954).

However, even many of those who suggest obtaining additional data admit that frequently it is not a possible solution. For example, "Admonitions that new data, or additional *a priori* information, are required to break the multicollinearity deadlock are hardly reassuring, for the gap between information on hand and information required to estimate a model fully is so often immense" (Farrar and Glaubner, 1957: 96). Similarly Johnston cautions "the remedy lies essentially in the acquisition, *if possible,* of new data which will break the multicollinearity deadlock" (Johnston, 1963: 207) [emphasis added].

[9] Indeed, problems in non-experimental data such as multicollinearity were long overlooked because the source of statistical analyses, such as regression and cross-tabulation, was the controlled world of the laboratory experiment. There, unlike the real world, variables can be manipulated so that the major explanatory factors under study operate independently of one another (Blalock, 1963: 233).

[10] The phrase is Johnston's (1963: 207), but his conclusion is less pessimistic and far reaching than the one reached here.

[11] Tukey's (1954) conclusion based on a discussion of both regression and path coefficients is: "The problem is highly complex and perhaps not capable of yielding any satisfactory solution."

[12] For example, Blalock's excellent analysis of the problem of multicollinearity is one of the few that analyzes both its nature and some of its consequences for non-experimental social scientists. Nevertheless, he does not discuss its consequences for policy analysis and evaluation. (Blalock, 1963: 234). Similarly Tufte's (1969) excellent analysis of methods of improving data analysis in political science points out some consequences of multicollinearity but not for policy evaluation.

[13] For example, this is the conclusion of Miller (forthcoming) concerning the effect of almost all programs to control and prevent delinquency, including detached worker programs. Similarly, after analyzing the various crime prevention and rehabilitation programs undertaken thus far, Stanton Wheeler and his associates conclude: "As of now, there are no demonstrable and proven methods for reducing the incidence of serious delinquent acts through preventive or rehabilitative procedures. Either the descriptive knowledge has not been translated into feasible action programs, or the programs have not been successfully implemented; or if implemented, they have lacked evaluation; or if evaluated, the results have usually been negative; and in the few cases of reported positive results, replications have been lacking" (Wilson, 1967: 73).

After surveying various efforts at compensatory education, the U.S. Civil Rights Commission (1967: 138) said "none of the programs appear to have raised significantly the achievement of participating pupils." There have been similar studies of Operation Head Start, which have had similar conclusions. (Westinghouse Learning Corporation Study, 1969; Evans, 1969). Most of these studies have been said to have serious methodological limitations; I am not referring to these questions but only to the problem of multicollinearity. Also see Cohen (1970: 8, 23-24) for an analysis of more recent evaluations of "Title I" programs which reach similar conclusions.

[14] See The President's Commission on Law Enforcement and the Administration of Criminal Justice (1967: 261). Another evaluation of the impact of lightning on crime in New York City also seems to have been plagued by multicollinearity.

[15] See Rosenthal and Jacobson (1968a) for descriptions of the experiments they conducted which focus on educational achievement and experimenter-subject interaction and bias. They also list several other controlled experiments focusing on social policies, primarily in the field of education. Also, see Campbell (1969) for a description of quasi-experiments with social policies and a bibliography of this field.

[16] There are several detailed descriptions of the Western Electric experiments and explanations of the Hawthorne effect. For example, see Homans (1950) and Roethlisberger and Dickson (1939).

[17] In the medical sciences a similar phenomenon is the "placebo effect" which is the introduction of a new treatment accompanied by improvement regardless of the nature of that treatment. See Shapiro (1960) for review of the history of the placebo effect.

It should be noted that the impact of the Hawthorne effect can be isolated to the degree that the experimenter is willing and able to use an additional experimental group which merely receives a placebo treatment (*i.e.,* it believes it is part of an experiment and believes that it is receiving an actual treatment). Then the results of this placebo group can be compared to the experimental group which actually received the treatment.

[18] See Rosenthal (1966) and Rosenthal and Jacobson (1962). Several other similar investigations by Rosenthal and others are cited in these two works. The highlights of a significant amount of their work in various contexts is described in Rosenthal and Jacobson (1968b).

[19] This explanation of a pattern of experimenter effects in the form of a self-fulfilling prophecy was developed earlier in Rosenthal's laboratory experiments with animals (Rosenthal, 1966: 151, 311 ff.).

[20] See Rosenthal and Jacobson (1968b: 22-23) for a description of other controls which were used and other tests of alternative hypotheses.

[21] Criminal statutes in most states allow the judge to incarcerate a convicted defendant or grant probation in common felonies. If the judge decides to imprison him, usually he can then set the term in prison within certain prescribed limits.

[22] By recidivism I simply mean an individual who is convicted of an offense after he has been convicted of a previous offense. The use of this term in no way implies the opposite of rehabilitation. For the sake of brevity, recidivism rates will almost always be stated in the short-hand terms of "success" rates or "failure" rates (which in no way imply any "existential" state). Since they are shorthand terms, the reader should note their precise operational definition which often varies among the studies described here.

In analyzing these studies formal probation with supervision and suspended sentences which do not involve supervision are considered together under the shorthand categcry of "probation." In most of the studies, almost all of the cases in this category involve formal probation with supervision. *No cases which involve probation plus some term of incarceration are included in the category "probation,"* although in the official data of some states, such as California, the term "probation," includes such cases. In this analysis these latter cases are included in the category of "incarceration."

23 I wish to thank the following individuals who graciously helped to provide the data which appears in this section: Ronald Beattie, Marie Vida Ryan, Charles Bridges, William Hutchins, Robin Lamson (all of the state government of California); State Assemblyman Craig Biddle of Riverside, California; Steven Kolodney (Public Systems Inc.), Don Gottfredson (National Council of Crime and Delinquency), Richard McGee (American Justice Institute), Charles Eichman (Florida Department of Corrections), H. P. Higgins and Carole Bartholemew (Minnesota state government), and John Yeager (Pennsylvania Department of Justice).

24 The Superior Court is the county trial court in California; its criminal jurisdiction includes all serious offenses (*i.e.,* all felonies and several major misdeameanors). The offenses included in their study are homicide, robbery, assaults, forged checks, auto theft, "other theft," sex offenses, drug law violations, and "other offenses."

The data in this study include all the Superior Court probation and jail cases for the first six months of 1966 for twelve of the thirteen counties and 30% of those cases from Los Angeles.

25 "Jail" refers here to a term of incarceration of no more than one year, which is served in a city or county jail. In California all terms of incarceration greater than one year are served in a state prison. The Beattie and Bridges study did not include offenders sentenced by the Superior Courts to state prison. The Beattie and Bridges study did not include offenders sentenced by the Superior Courts to state prison, but they are analyzed in studies described below.

The follow-up period in this was twelve months from the time of the individual's release to the street on probation or following incarceration. This is a limitation only in assessing the general degree of recidivism. (Other studies have indicated that while most recidivism occurs during the first year following release, a significant degree does occur in the next year.) This does not seem to be a limitation for assessing the differences, if any, in recidivism rates for different types of treatment. There is no evidence in other studies that the recidivism rates for different types of treatment would vary significantly from the first to second year. Nor is there any substantive reason to entertain such a hypothesis.

26 The larger project of which this study is a part will analyze the data collected by Beattie and Bridges in more detail than was possible in their own study. For example, additional characteristics will be controlled simultaneously. Regression analysis of their data will also be carried out to assess more precisely the relative effect on recidivism of the various offender characteristics and types of treatment. A preliminary effort at such an assessment is described below.

27 All defendants granted probation or "probation plus jail" in California during 1956 to 1958 were included in the analysis, except those in Los Angeles and Alameda (Oakland area) counties for which "there was inadequate information at that time."

28 Davis only presents percentages for the combined categories probation and "probation plus jail"; I have recalculated his raw data to ascertain percentages for these categories separately.

29 England explicitly states only a preoise criterion of "failure"—if a probationer is subsequently convicted of a misdemeanor or felony. Therefore, it is likely that included in his "success" group are some individuals who were arrested but not convicted, or who committed a technical violation of probation but were not convicted of a new offense.

Approximately 75% of the sample had committed less serious offenses such as "white-collar" crimes (embezzling, mail fraud), counterfeiting, forgery, bootlegging, price control violations, draft evasion, and transportation of stolen cars across state lines.

https://doi.org/10.2307/3052911 Published online by Cambridge University Press

[30] Internal memoranda of the California Department of Corrections, May 1, 1970, and April 20, 1967. I am indebted to Marie Vida Ryan, Senior Statistician of the CDC for graciously providing these data and many other aids to this study.

[31] Because of the differences in the categories used by Beattie and Bridges, this "failure" rate is probably somewhat of an underestimation in comparison to the CDC data.

[32] Again this percentage is probably somewhat of an underestimation.

[33] The definitions of "success" were exactly identical in both studies— no subsequent arrests. However, the follow-up period in the PSI study was three years and in the Beattie and Bridges study it was only one year. This should not have significantly lowered the "success" rate in the PSI study because most studies indicate that the preponderance of recidivism occurs during the first twelve months. Indeed, the PSI data themselves indicate almost 70% of the recidivism of those in its study occurred during that period.

[34] Upon release 60% of the *Gideon* early releases had been incarcerated for less than eighteen months and only 46.5% of the full-term releases had been incarcerated for less than that time, and Eichman found this difference statistically significant.

[35] Glaser defined "success" rate as it is in the Beattie and Bridges study — no subsequent arrests. However, Glaser is quite careful to make distinctions among the "non-success" group: Only 31.1% of Glaser's total sample were subsequently returned to prison; 16.7% were subsequently arrested and/or incarcerated for a nonprison sentence. Generally it is only this 31.1% that Glaser refers to in his analysis as "recidivists."

[36] When offense and age are simultaneously controlled, for sex offenses committed by individuals over thirty years old the simultaneous impact of these factors is greater than the type of treatment which they receive. The "success" rate for these offenders is 86.9% (114) for those granted probation and 84.2% (32) for those incarcerated. When offense and prior record are simultaneously controlled, there are similar patterns of a greater impact of these simultaneous factors for sex offenses committed by whites (almost identically high "success" rates for both types of treatment), for sex offenses committed by individuals with no prior record (almost identically high), and for burglary committed by individuals with no prior record (almost identically moderate "success" rates). See Beattie and Bridges (1970: 21-25).

[37] For example, as noted, youthfulness has a major impact on recidivism, but for offenders under twenty years old and for those twenty to twenty-four years old the "success" rates are higher for those granted probation — 54.0% (176) and 58.0% (712) respectively — than for those incarcerated — 44.4% (180) and 42.7% (924) respectively. Similarly, as noted, whether an offender is a Negro has a significant impact on recidivism, but for Negro offenders the "success" rates are much higher for those granted probation. The degree of prior record also has a major impact on recidivism, but for those with the greatest degrees of prior record, the "success" rates are significantly higher for those granted probation. See Beattie and Bridges (1970: 21-25).

[38] The Coleman Report (1966) had precisely the same difficulty with insufficient variation among the data points for the independent variable of "class size." There was an insufficient number of small classes. Some critics have suggested that this led the Coleman analysis to underestimate the potential impact of class size — especially a small class size — on educational achievement. This type of insufficient variation is common in the analysis of policy data.

[39] The PSI study does not state the precise correlations among its independent variables, but some of its raw data indicate this degree of intercorrelation (*e.g.*, most individuals — 71.1% [1,972] — who have the most serious prior records also received long terms of incarceration, while only about 5% of the entire sample received a long term of incarceration and had no prior record). See Kolodney, *et al.* (1970: III-20). In an anlysis of the PSI data which is planned later these precise correlations will be ascertained and further tests for multicollinearity will be applied.

[40] Seventy to eighty percent of those in the experimental group resided in their own homes. The remainder were placed in a foster or group home because it appeared to the CTP investigators that they could not

live in their own home and remain non-delinquent. These 20-30% usually spend at least part of the time in their own home, but their lives generally are somewhat more constrained. See Warren (1967: 5). However, it does not seem that this constraint is significant enough to suggest that they are no longer experiencing freedom. Their experience is still much like that of those in the experimental group who live at home and it is still radically unlike that of those in the incarcerated or control group.

41 Tests administered both at intake into the Youth Authority and after release (after treatment in the case of the experimental groups and after discharge from institution in the case of the control group) indicated that "although both groups showed improvement from pre-test to post-test, the experimental group showed considerably more positive change than the control group, together with a higher level of personal and social adjustment." See Warren (1967:7).

42 Labeling is a special case of a self-fulfilling prophecy in which an individual is named or given a "label" which then seems to often create a self-fulfilling identity of personal definition of his behavior. For discussions of this process see H. Becker (1963), Kitsuse (1962: 247-256), and Lemert (1967: Ch. 13).

43 The populations of Sacramento and Stockton in 1960 were 191,667 and 86,321 respectively. Seven percent and 10% of these populations respectively were Negro, 8.1% and 16.8% were Mexican-American and both had rather large portions of their labor force in white collar occupations (54.6% and 46.1% respectively).

44 In an apparent effort to remedy this flaw, Phase 2 of the CTP experiment was extended to predominantly Negro areas of Los Angeles and Oakland. However, for reasons that can only be speculated upon, Phase 2 does not include random assignment of convicted delinquents. Instead the youths are assigned to the community treatment program after screening by the project staff. Moreover, there is no control group whatsoever. Instead, the effect of the program is measured by comparing the "failure" rate of youths assigned to it with equivalent statewide rates for youths of the same middle to older adolescent age range. After a fifteen-month follow-up period of parole exposure, the "failure" rate (defined as parole revocation) for the project's youths is 39% compared to 48% for the statewide group of that age. See U.S. Task Force on Corrections (1967: 42).

45 I will use among other things (a) data from Ralph England's unusual probation population; (b) George Davis' correlations between absolute rates of probation for different offense and the corresponding recidivism rates, which were not analyzed here.

46 See Levin (forthcoming: Ch. 5) for the details behind these general statements.

CASES

Gideon v. Wainwright 372 U.S. 335 (1963).

REFERENCES

BEATTIE, Ronald H. and Charles K. BRIDGES (1970) Superior Court Probation and/or Jail Sample. Sacramento: Bureau of Criminal Statistics, Department of Justice.

BECKER, Howard S. (1963) Outsiders. New York: Free Press.

BECKER, Theodore L. (1969) The Impact of Supreme Court Decisions. Oxford: Oxford University Press.

BLALOCK, Herbert Jr. (1963) "Correlated Independent Variables: The Problem of Multicollinearity," 62 Social Forces 233.

BOWLES, S. and H. LEVIN (1968) "The Determinants of Scholastic Achievement," 3 Journal of Human Resources 3.

CAMPBELL, Donald T. (1969) "Reforms as Experiments," 24 American Psychologist 409.

COHEN, David (1970) "Politics and Research: The Evaluation of Social Action Programs in Education." 40 Review of Educational Research 213.

COLEMAN, James S., *et al.* (1966) Equality of Educational Opportunity. Washington, D.C.: Government Printing Office.

DAVIS, George (1964) "A Study of Adult Probation Violation Rates by Means of the Cohort Approach," 55 Journal of Criminal Law, Criminology and Police Science 70.

DYE, Thomas (1966) Politics, Economics, and the Public. Chicago: Rand McNally.

EICHMAN, Charles J. (1966) The Impact of the Gideon Decision Upon Crime and Sentencing in Florida. Tallahassee: Florida Division of Correction.

ENGLAND, Ralph W. Jr. (1957) "What is Responsible for Satisfactory Probation and Post-Probation Outcome?" 47 Journal of Criminal Law, Criminology and Police Science 667.

EVANS, John (1969) "The Westinghouse Study: Comments on the Criticisms," in David G. HAYS, Britannica Review of American Education, Vol. 1. Chicago: Encyclopaedia Britannica.

FARRAR, Donald E. and Robert R. GLAUBNER (1967) "Multicollinearity in Regression Analysis: The Problem Revisited," 49 Review of Economics and Statistics 92.

FRY, Brian and Richard WINTERS (1970) "The Politics of Redistribution," 64 American Political Science Review 508.

GLASER, Daniel (1964) The Effectiveness of a Prison and Parole System. Indianapolis: Bobbs Merrill.

HOMANS, George (1950) The Human Group. New York: Harcourt, Brace.

JACOB, Herbert and Michael LIPSKY (1968) "Outputs, Structure and Power: An Assessment of Changes in the Study of State and Local Politics," 30 Journal of Politics 510.

JACOB, Herbert and Kenneth VINES (1965) Politics in the American States. Boston: Little, Brown.

JOHNSTON, John (1963) Econometric Methods. New York: McGraw-Hill.

KITSUSE, John T. (1962) "Societal Reaction to Deviant Behavior: Problems of Theory and Method," 9 Social Problems 247.

KOLODNEY, Steven, *et al.* (1970) A Study of the Characteristics and Recidivism Experience of California Prisoners. San Jose: Public Systems Incorporated.

LEMERT, Edwin M. (1967) Human Deviance, Social Problems, and Social Control. Englewood Cliffs, N.J.: Prentice Hall.

LEVIN, Martin A. (forthcoming) Urban Political Systems and Judicial Behavior: The Criminal Courts. Cambridge: Harvard University Press.

.... (1970) "An Empirical Evaluation of Urban Political Systems: The Criminal Courts," in Sam KILPATRICK and David MORGAN, Urban Politics: A System Analysis. Glencoe: Free Press.

MEYER, John R. and Edwin KUH (1957) The Investment Decision: An Empirical Analysis. Cambridge: Harvard University Press.

MILLER, Walter (forthcoming) City Gangs. New York: John Wiley.

MUIR, William Jr. (1967) Prayer in the Public Schools: Law and Attitude Change. Chicago: University of Chicago Press.

ORCUTT, Guy, *et al.* (1961) Microanalysis of Socioeconomic Systems: A Simulation Study. New York: Harper.

PENNOCK, J. Roland (1966) "Political Development, Political Systems, and Political Goods," 18 World Politics 415.

PRAIS, S. J. and H. S. HOUTHAKKER (1955) The Analysis of Family Budgets. Cambridge, England: University Press.

President's Commission on Law Enforcement and Administration of Criminal Justice (1967) The Challenge of Crime in a Free Society. Washington, D.C.: Government Printing Office.

ROETHLISBERGER, Fritz J. and William J. DICKSON (1939) Management and the Worker. Cambridge: Harvard University Press.

ROSENTHAL, Robert (1966) Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts.

ROSENTHAL, Robert and Lenore JACOBSON (1968a) Pygmalion in the Classroom. New York: Holt, Rinehart and Winston.

.... (1968b) "Teacher Expectations for the Disadvantaged," 218 Scientific American 19.

SHAPIRO, Arthur (1960) "A Contribution to a History of the Placebo Effect," 5 Behavioral Science 109.

STONE, Richard (1954) The Measurement of Consumers' Expenditure and Behavior in the United Kingdom. Cambridge, England: University Press.

TUFTE, Edward (1969) "Improving Data Analysis in Political Science," 21 World Politics 641.

TUKEY, J. W. (1954) "Causation, Regression, and Path Analysis," in Oscar KEMPTHORNE, *et al.*, Statistics and Mathematics in Biology. Ames, Iowa: Iowa State College Press. Quoted on page 237 of Herbert BLALOCK, Jr. (1963) "Correlated Independent Variables: The Problem of Multicollinearity," 62 Social Forces 233.

U.S. Commission on Civil Rights (1967) Racial Isolation in the Public Schools. Washington, D.C.: Government Printing Office.

U.S. Task Force on Corrections (1967) Task Force Report: Corrections. Washington, D.C.: Government Printing Office.

WARREN, M. Q. (1967) The Community Treatment Project After Five Years. Sacramento: California Youth Authority.

WASBY, Stephen L. (1970) The Impact of the U.S. Supreme Court. Homewood, Ill.: Dorsey Press.

Westinghouse Learning Corporation Study (1969) The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development.

WILSON, James Q. (1968) City Politics and Public Policy. New York: John Wiley.

.... (1967) "The Crime Commission Reports," 9 The Public Interest 64.

.... (1964) "Problems in the Study of Urban Politics," in E. H. BUEHRIG, Essays in Political Science. Bloomington: Indiana University Press.