U.S. Department of Justice National Institute of Corrc

Screening for Risk

A Comparison of Methods

72276

Screening for Risk

A Comparison of Methods

Stephen D. Gottfredson
Center for Metropolitan Planning and Research
The Johns Hopkins University

and

Don M. Gottfredson
School of Criminal Justice
Rutgers University

August 14, 1979

This report was prepared under Grant No. AM-1 from the National Institute of Corrections, U.S. Department of Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Contents

Contents (continued)

Tables

Tables (continued)

Tables (continued)

Figures

Preface

This report is intended to provide the correctional community with advice concerning the development and use of statistical models commonly employed in criminal justice prediction studies. Since the use of statistical decisionmaking aids is increasing in correctional settings, an evaluation of the potential strengths and limitations of various methods often used or advocated is provided.

Advice concerning methodological issues is always problematic. What works well in one study or setting may work less well or fail miserably in another. What one researcher perceives as parsimony may be perceived as simple-minded by another. Throughout this effort, we have sought to remain unbiased. While we admire parsimony and, indeed, simplicity, we likewise appreciate elegance. We trust that we have presented the issues fairly—this, at least, was our intent.

S.D.G.
D.M.G.
August 14, 1979

Acknowledgements

This report could not have been prepared, nor the studies upon which it is based conducted, without the active collaboration of a number of people.

Professor Leslie T. Wilkins of the School of Criminal Justice, State University of New York at Albany, served as an advisor to the project from its inception. Many of the ideas contained herein are his, and his sage and helpful advice and guidance is most appreciated.

Dr. Peter B. Hoffman, Director of Research for the United States Board of Parole, was instrumental in extending the original Parole Decisionmaking Project data base to the advantage of this project. Further, along with Professor Wilkins, he met with and advised the authors throughout the project.

We hasten to add, however, that any errors of omission or commission are ours alone. While much of the "good" is theirs, any remaining "bad" is ours.

Some portions of this report were originally drafted by other colleagues. In particular, we wish to thank K. Andreason, W.H. Brown, G. Dodsley, M.G. Neithercutt, G.E. Pasella, D. Pfoutz, and S. Springer, all of the National Council on Crime and Delinquency Research Center; J.E. Goodnow II, now at Bell Laboratories; and T. Land and H.M. Keintz, both at Johns Hopkins. In addition to performing many analyses and refining most of the statistical programs used, Mr. Keintz took responsibility for the difficult task of drafting an explanation of the log-linear analyses. His help is gratefully acknowledged.

To Ms. Verna Cavey, our Administrative Assistant, and Ms. Kathy Young, our Research Assistant, we owe special thanks for increasing the quality of our effort.

The help of the members of the United States Parole Commission, who collaborated in the design and implementation of the original collection of the data used and who have supported this and related earlier efforts, is appreciated.

Finally, the National Institute of Corrections, particularly through the project monitor, Mr. Larry Solomon, provided both monetary and substantive support.

1 Introduction

Research workers in the criminal justice arena have long been on the forefront of an area of study that greatly concerns all behavioral science—the prediction of specified events or behaviors. Problems of behavioral prediction are both practical and methodological; decisions requiring predictions must and will be made in virtually any area affecting people. College administrators and admissions committees must attempt to predict the future behavior of large numbers of applicants. Employment counselors and personnel managers must attempt to predict the future behavior (or satisfaction) of prospective employees or career-seekers. Judges must estimate the likelihood of future offenses, and members of paroling agencies must attempt to predict the future behavior of large numbers of eligible inmates. Accordingly, considerable effort has been expended in attempts to aid the decisionmaker in the predictive situation.[1]

Major methodological problems of behavioral prediction can be classified into five general categories: 1) the relative efficiency of "clinical" versus "actuarial" or "statistical" methods of prediction, 2) the relative efficiency of different actuarial approaches, 3) the base-rate problem, 4) the reliability problem, and 5) the cross-validation of predictive measures. Clearly, these five categories are neither exclusive nor exhaustive. As our discussion will demonstrate, the consideration of any one issue necessitates a consideration of the others.

While we treat the first topic cursorily, it is of great importance. The history of the debate on clinical versus actuarial approaches to behavioral prediction is long. Often, the "problem" has been interpreted as, "Which approach gives better results, the intuitive, inductive *clinical* approach, or the deductive, objective *actuarial* approach?" The polarizing effect of such a formulation of the problem is clear. Indeed, while recognizing its polarizing character, Gough nonetheless takes this position in his classic review of the area:

> *The problem may be posed in a brief and simple manner: In any given predictive situation which method is better—i.e., more accurate and more informative in a scientific way—that of the clinician or that of the actuary?*[2]

We prefer a different formulation of the clinical versus actuarial prediction problem—one which vitiates the "versus."[3] In many practical situations, decisions based on predictions will be made, and help toward more rational decisions may come from either sector or from a combined use of information. In general, behavioral scientists are not, and have no interest in becoming, decisionmakers. They do, however, have great interest in the prediction of behavior and in the decisionmaking process. The "clinical versus actuarial" problem is less important when asking, "Which approach is better?" but more important when we ask, "Can behavioral scientists using actuarial methods be of practical help to the decisionmakers?" The focus of the present study is therefore on issues of the development and use of actuarial methods, not on comparisons of such devices with clinical judgments.

It does appear now that statistical prediction methods may have a variety of practical uses. It is clear, for example, that actuarial prediction can improve substantially upon clinical, intuitive approaches in a statistical sense.[4,5] As a result, actuarial methods have been used not only in practical selection problems but in program development applications, offender placement decision situations, and in decision guidelines models, as well as for research purposes.

Despite considerable experience in the development of such methods, however, there is a great deal of theoretical and practical debate about the most efficient (most valid, least costly, most operationally useful) methods for selecting and combining information with some predictive utility. Because operating agencies are increasingly involved in the development and use of such screening devices, correctional research workers and practitioners involved should have a clear statement of the advantages and disadvantages of the most prominently used methods.

Recent reports have suggested that, in practice, statistical techniques that are theoretically less powerful and that are computationally and procedurally relatively simple may demonstrate equal or superior predictive validity to that obtained by more complex and theoretically superior methods.[6] To date, however, despite the practical importance of identifying those predictive methods likely to be of greatest aid in screening for risk, no comprehensive comparative assessment of specific methods has been made. While a few comparative studies have been conducted,[7] each has suffered from one or more limitations, stemming from:

1) Lack of attention to the base-rate problem, i.e., to relative improvement over the "success rate" for the sample as a whole.
2) Failure to cross-validate, i.e., to test the method on new samples.
3) Lack of application of methods to the same sets of data.
4) Consideration of only a few of the most widely used methods.

One study, however, did clearly suggest that less sophisticated statistical techniques may indeed provide substantive conclusions of equal power to those provided with the use of more sophisticated methods[8]—a suggestion supported by the confirmed findings of several comparative studies cited above.

The research we report here overcomes the four limitations noted. Since our study had the comparison of efficiency as its primary goal, all analyses were carried out on the same data base. The data used are at least as adequate to the purpose as any other currently available; advantages of the data base are discussed in a later section. All comparisons are made with attention to the base rate and validation issues. Finally, we included in the comparison a variety of markedly different methods.

Appendices not contained in this report are available through the National Institute of Corrections Information Center, 1790 30th Street, Suite 314, Boulder, CO 80301, and through the National Criminal Justice Reference Service. These appendices contain the data base, assessment of potential predictor variables, and the analytic programs used.

2 Issues in the Development of Risk-Screening Devices

To address adequately the question of the most useful methods to employ, it is necessary to consider in more detail the issues already mentioned, such as the problems of the base-rate, cross-validation, and reliability. Indeed, it is only by reference to these issues that the question of the relative efficiency of different methods of combining predictive information can be resolved. Some problems of behavioral prediction, moreover, are practical and policy-oriented, rather than simply statistical. This section discusses both types of problems in more detail.

Statistical Issues

The Base-Rate Problem

The base rate for any given event may be defined as the relative frequency of occurrence of that event in the population of interest.[9] The base rate *problem* is actually a combination of many difficulties. Perhaps the most basic of these difficulties is that often base rates are not considered at all. In 1955, Meehl and Rosen summarized the consequences of failure to consider base rates and concluded that "almost all contemporary research reporting neglects the base rate factor and hence makes evaluation of test usefulness difficult or impossible."[10]

Behavioral scientists are often concerned with the prediction or classification of events, but such classifications or predictions have often been based on criteria that produce larger errors than would the simple use of the base rate.

To the extent that the base rate differs from 0.50, difficulty of prediction of an event increases. Thus, the more infrequent an event, the greater the likelihood of inaccurate prediction.[11] While this seems intuitively true for rare events, it must be remembered that the occurrence of very frequent events requires the simultaneous occurrence of very rare events (unless the probability of an event is precisely 0 or 1).

As an example of the difficulty of such prediction, suppose that the base rate for failure on parole is 0.20. Given this information alone, we know that we would make correct predictions 80 percent of the time if we simply predict that *no one* will fail on parole. We will also, of course, be wrong 20 percent of the time. Note that, given only the base rate as a guide, we have no way of knowing which 20 percent will in fact fail.

Now let us assume that a predictive device has been developed that allows us to correctly predict parole outcomes with 78 percent accuracy. Even given this apparently powerful device, we would still be better off in expecting that no one will fail on parole—that is, in "predicting" performance on the basis of the base rate alone. While our predictive device does beat the simple chance rate of 50 percent, the chance rate should, as noted by Meehl and Rosen, be viewed as a function of the base rate. Hence, in this example, the chance rate would be substantially higher than 50 percent, and the simple base rate is still the more accurate.

The practical consequences associated with varying base rates are another part of the "problem." In practical selection problems, for example, one must take into consideration the consequences, in terms of monetary or social costs, of two kinds of predictive errors. That is, some predicted "successes" will fail; some expected "failures" will succeed. It may be, for example, that failing to identify potential parole failures prior to release is more costly to society than is any effort expended on preventive measures for those who would not fail anyway.

Another problem associated with the base rate is common. Often, devices that may have predictive validity for the population for which they were designed (i.e., are more accurate than the base rate) are used on different populations, for which the base rate of occurrence is different. [12] (This problem is most evident with predictive devices that utilize an "automatic" cutting score. For example, a device used to predict parole failure of institutionalized adult offenders and based on a cutting score may prove of little use for predicting violation among juvenile offenders. It should be noted, however, that successful prediction often can be increased through manipulation of the cutting score. A related problem is the potential heterogeneity of the offender sample providing the basis for prediction. [13]) In general, however, this reflects a more fundamental problem—that of unwarranted generalization to new samples. This issue is not further addressed in the study reported here, and generalizations of the results to other samples should not be made.

Finally, of course, it is often difficult, if not impossible, to know the "true" base rate of occurrence of a specified event or behavior in the population of interest. We cannot, for example, know the true base rate for parole violation for all offenders considered for parole. Since not all are in fact paroled, we can at best identify the base rate for *known* violations by *paroled* inmates.

Ohlin and Duncan were among the first to give practical attention to the base rate problem. They developed an "index of predictive efficiency" (the percentage change in errors of prediction over the base rate resulting from the use of a prediction method) to assess the usefulness of prediction devices. [14] The statistic used in this study to compare the effectiveness of various actuarial methods was the "Mean Cost Rating" (MCR) of Duncan and Duncan—a successor to the index of predictive efficiency. [15,16] This index can be used to assess the relative efficiency of various prediction devices with respect to the base rate problem. The MCR takes on values in the range 0 to 1, and its relative magnitude gives an indication of prediction above the base rate.

Statistics of this type are related to "proportionate reduction in error" measures, in that they attempt to offer an evaluation of predictive power above that of the chance rate. In this context, reliance on standard correlational techniques may be misleading, since they do not explicitly address the base rate issue. Hence, we prefer to base our evaluations of predictive efficiency primarily on the MCR.

While the sampling characteristics of the statistic have not previously been known, it recently was demonstrated that the statistic is related to Kendall's tau, [17] and recent work demonstrates that the MCR can be interpreted via the framework of Signal Detection Theory. [18]

The Validation Problem

The adage that no two people are exactly alike is properly extended to groups of people. No two groups of people are identical. If, however, the groups have been selected by some appropriate mechanism (such as random sampling), they can be expected to have a great deal in common in terms of both their overall characteristics and the interrelations of various individual characteristics. It is this similarity of relations within different groups of people upon which all statistical predictions ultimately rely. It is assumed, for example, that if in one group of subjects the younger do better in relation to some outcome, then in a similar group of subjects the younger again will do better. Indeed, prediction methods are intended to help estimate, on the basis of some group of people available for study, how members of other similar groups will behave. In doing so, there is a risk of overestimating the extent to which relations found in one sample can be used to explain relations found in a similar sample.

There is no way to distinguish, *within the original sample alone,* how much of the observed relation is due to characteristics and underlying associations that will be shared by new samples and how much is due to unique characteristics of the first sample. Mistaking unique, peculiar variation for more general variation is referred to as "sample overfitting." As might be expected, those methods which fit the current (construction) sample very accurately are more prone to overfitting than are those which fit the construction sample with less rigor. In addition, one of the maxims of any statistical procedure is that large samples reduce the relative importance of strictly individual variation and improve the chances that observed relations are due to general factors and thus are likely to be observed in other groups. This implies that, other things being equal, predictions based on a small number of observations, like those produced by methods which employ sample subdivisions, are more prone to sample overfitting than are those based on larger samples.

The apparent power of a prediction device developed on a sample of observations thus derives from two sources. The first, already discussed, is the detection and estimation of "underlying" relations that are likely to be observed in any similar sample of subjects. A second and troublesome source is the peculiar or individual properties of the specific sample. It is imperative that in any prediction study intended for practical application in new samples the relative importance of these two sources of predictive power be estimated. Failure to do so may be expected to result in the overestimation (sometimes large) of the actual utility of the instrument as a predictor.

The estimate of the relative importance of the two sources of predictive power is typically accomplished by randomly dividing the group under study into two samples, or by gathering (in the same fashion) a similar sample specifically for this purpose. It may be argued that a better test, given that the instrument may be intended for future use, is given by selecting another sample from a later time period. Predictive devices (often equations) are constructed using only one of these groups (the *construction* sample). The equation or equations that have been developed are then applied to the other group (the *validation* sample). The correlation or other measure of association that results provides an estimate of the predictive power likely to result from subsequent applications of the method to similar groups. Taking repeated samples over time would, of course, provide yet better evidence. This procedure is known as *validation* or *cross-validation*. Apparent loss of predictive power from the construction to the validation sample is known as *shrinkage*, and it results, in large part, from the overfitting of the device or equation on the construction sample.

An important issue in proposals for practical applications of prediction devices, then, is that of potential shrinkage in power from the construction to validation samples. To further our investigation of the relative efficiency of several different actuarial approaches, all approaches are compared relative to validation samples. The extent of any differential shrinkage, of course, is of particular interest.

The Reliability Problem

No prediction device can be better than the data from which it is derived. An unfortunate occurrence in behavioral science research is that we often do not know, and sometimes cannot determine, the reliability of the information upon which predictions are based. Sources of unreliability are numerous, as considered in detail by Campbell and Stanley [19] and by Cronbach. [20] Descriptive crime and delinquency statistics, moreover, tend to be particularly unreliable for a number of reasons peculiar to the criminal justice system. [21]

The consequences of a failure to consider reliability (or the lack thereof) can be serious indeed. [22,23] The present study provides indices of reliability of one important kind — that of the coding or classifying of data elements in abstracting a large number of items from case files.*

Policy Issues

Concerns of policy development and implementation — important as they are to correctional administrators, judges, paroling authorities, or criminal justice planning bodies — lie outside the scope of the present study and this report. Still, it may be helpful at least to mention some policy issues that currently are the subject of much debate and involve the general problem of prediction.

The centrality of prediction to many concerns of the criminal justice system may be emphasized by recalling that major traditional utilitarian aims of that system—such as deterrence, treatment (rehabilitation) and incapacitation—are all forward-looking, crime preventive goals that require prediction in some sense. The concept of *deterrence* involves the "prediction" that punishment of known offenders will discourage others from crime. The concept of *treatment* involves the "prediction" that offenders may be changed in order to reduce the likelihood of repeated offending; and that of *incapacitation* requires the "prediction" of new offenses if offenders are not restrained from committing them. (The aims of desert — that is, of applying deserved punishment — or of retribution or retaliation do not involve prediction [nor, for that matter, a utilitarian attitude]. Rather, they look backward only, to the gravity of the harm done or the culpability of the offender.) [24]

In view of the centrality of prediction to many policy issues in criminal justice, we seek in this section to identify a few such issues in which prediction methods play an important part. These concern, first, the relevance of statistical prediction models to criminal justice goals; second, questions of "what to predict, when, with what information, for what purposes;" and finally, some general limitations of prediction methods that are particularly important for policy decisions about their use.

In specific applications of prediction methods, controversies regarding their use are apt to arise from scientific empirical evidence, from ethical value perspectives, or from both, but this does not require that these concerns be confused. Issues of value and of evidence should be identified in order to add clarity to arguments in policy formulations. This report has a limited focus on selected empirical issues, and it deals very little with the many important issues of values. We seek in this section merely to point out the distinction between the scientific and value questions, to suggest how they often are related, and to note some illustrative ethical issues that may be of fundamental importance to policy decisions concerning the operational applications of prediction methods.

Relevance to Goals

If the decision problem is one that requires predictive information for the selection of alternative actions, then the use of prediction methods to aid in the decisions should be considered. In a previous section of this report, we discussed briefly the continuing debate concerning "clinical" *vs.* "statistical" prediction; two claims of that discussion may be reiterated here. First, it generally has been found that statistical prediction methods can be developed that are more reliable and valid than unguided or intuitive clinical predictions. That is, they are more dependable and they work better. Second, statistical and clinical methods may be used together in various, possibly mutually supportive, ways; thus, the use of statistical prediction methods does not necessarily imply that clinical judgments may not (or should not) be used at all in decisionmaking.

The relevance and potential utility of statistical prediction methods to criminal justice obviously depend upon the objectives of the decisions involved. Unfortunately, the objectives are completely agreed upon only rarely, and rarely are they stated with the clarity needed for any careful analysis. Nevertheless, it may be claimed that

*See Appendix 3, "Assessment of Potential Predictors," available from the National Institute of Corrections National Information Center, Boulder, CO.

when the aims include those of incapacitation or treatment (more generally, the reduction of the probabilities of future crimes by the offenders concerned), statistical prediction methods may be expected to be relevant and perhaps potentially useful. The explication of agency or decisionmaker objectives is therefore an obvious requisite to deciding whether prediction methods are to be regarded as possibly useful. If the aim is limited to the provision of deserved punishment, then it is difficult to see how prediction methods can be helpful. If, on the other hand, the objectives include crime reduction purposes with respect to the population of offenders being considered, then it is difficult to see how prediction is *not* a relevant issue.

When to Predict What, With What, and Why

What is to be predicted depends, of course, on prior judgments concerning goals and objectives, but it may depend also upon the "why" part of the question. Specific sub-objectives may be perceived as steps toward more general aims. Thus, the specific objectives of an intended operational use of prediction methods need to be identified and considered as well as the general aims. As described in this report, choices — often rather arbitrary ones — must be made in defining a criterion of "success" or "failure" or other classification of outcomes to be predicted. Whether arrests, as well as convictions, should be counted as indicants of "failure" is an issue providing an obvious example of a definitional choice that may be influenced heavily by issues of value in relation to the specific intended operational use of the instrument.

One guide is given, in some such choices, from an understanding of the procedures underlying the development and validation of any such device. That is, a specific criterion must be used, and any generalizations of predictive validity to different criteria must be suspect until such validity has been demonstrated. Classifications relevant to one purpose may have no relevance to another. For example, classifications relevant to the risk of repeated offending may have no relevance or utility for the problem of assignment to different treatment programs. The decisionmaker is well advised that, although a particular prediction instrument may provide helpful information about one objective, it may give no (or, indeed, faulty) information about other outcomes that may be related to other decision objectives. It is plausible, therefore, that the inclusion of "arrests" as an unfavorable outcome element may be appropriate for some purposes and not for others. In one application, there may be no issue of fairness; in another, this concern may arise.

The "why" question is related also to "when" and "with what." Generally, four categories of purposes for developing prediction methods may be discussed. These involve research aims and applications for program planning, selection, and decision policy.

The process of developing prediction instruments involves the testing of hypotheses and thus may provide an aim in itself. Further, however, the purpose may be the provision of a research tool. In many program evaluation studies, for example, such measures are used to provide an indication of the "prior probabilities" of an outcome, i.e., of the likelihood of that outcome without treatment or before treatment. This may be useful in making comparisons of outcomes for groups treated differently.

Alternatively, the objective may be to provide a tool for the general screening of a population for some program planning purpose. An example might be the screening of all offenders received in prison using an "escape risk scale" in order to identify a pool to be considered for immediate placement in a minimum custody setting.

Or, the intention might be the development of an instrument to be used as an aid in selection, such as for probation or parole. A similar application is the use of such a device in allocation of probationers or parolees to differing levels of supervision, e.g., providing only minimal supervision of those in the most favorable "expected" outcome groups.

A somewhat different but related use of prediction methods is found in their inclusion, with other information, in policy or "guidelines" models such as those now sometimes used in sentencing or parole decisionmaking. [25]

In any selection application, value questions arise particularly in relation to two sets of concerns: these are issues surrounding the "with what?" question (what items may justifiably be included in the instrument) and the issue of errors in prediction.

The first question is one of values and of fairness. For example, suppose that the purpose of an instrument under development is to provide an aid to selection for parole. (The paroling authority has determined already that the issue of risk of new crimes is an appropriate one for inclusion in paroling policy.) Suppose further that the item "income level" is found uniquely and rather powerfully predictive of new crimes, such that offenders who have been poor are worse risks. In this wholly hypothetical example, the paroling authority obviously must resolve the policy issue: Is it fair to include the item?

The second problem, of errors in prediction, must be confronted as well. Whether clinical or statistical predictions are made, there will be errors of two kinds: predicted successes who fail ("false positives") and predicted failures who succeed ("false negatives"). This problem also raises complex questions of both evidence and values.

At different stages in the correctional process, correctional decisionmakers may have markedly different amounts and types of information available to them (Figure 1). Accordingly, an important issue in the development of actuarial aids involves decisions about the information to be included in the model. For example, a parole board might wish to use such a device at the intake stage, when information concerning prison adjustment during the instant commitment is not yet available. A board that uses such a device at the stage of the parole hearing, however, may well find such data informative. Again, the selection of items to be included in the device must be guided by its intended use.

All these issues are made more complex by the fact that neither organizations nor persons have single goals. Rather, they have multiple, sometimes conflicting, aims. Suffice it to say that prediction methods useful for one purpose may have little utility for another. Some general limitations of prediction methods may be noted, all of which should be remembered when policy issues surrounding their use are debated.

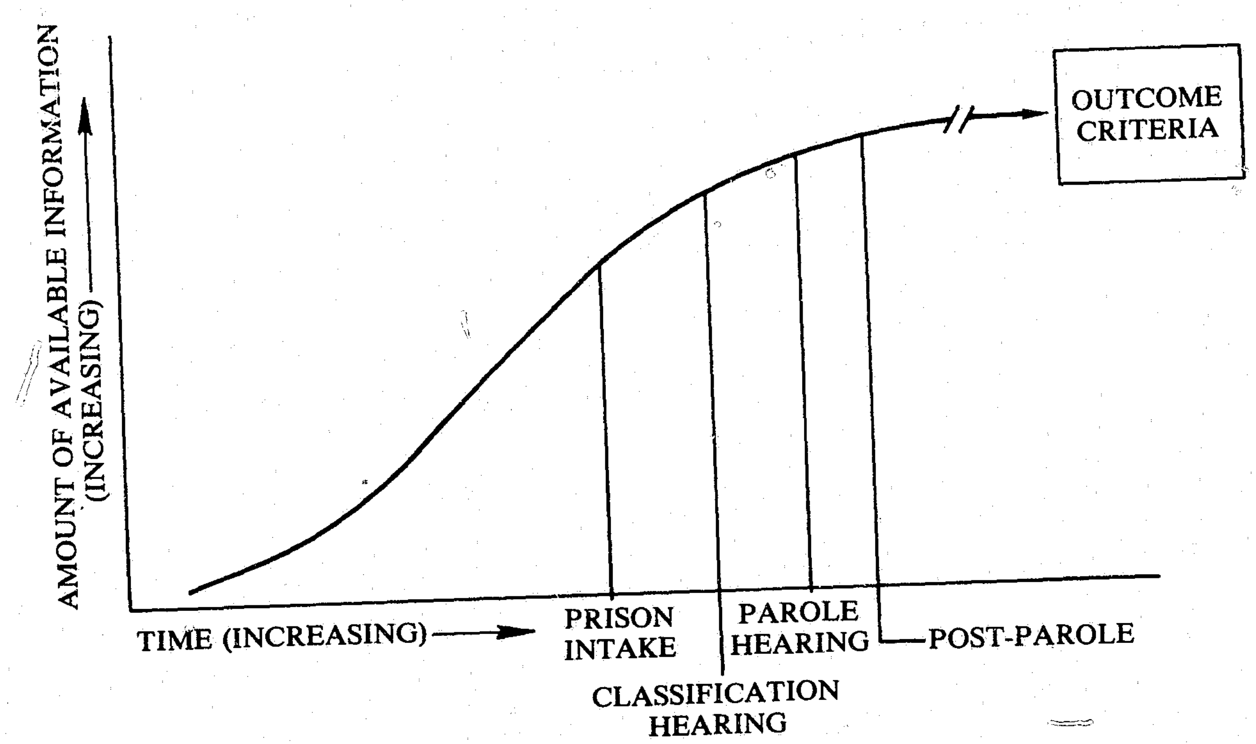

FIGURE 1
HYPOTHETICAL DISTRIBUTION OF THE AMOUNT OF INFORMATION AVAILABLE TO DECISIONMAKERS AS A FUNCTION OF TIME IN THE CORRECTIONAL PROCESS

General Limitations

Perhaps the greatest limitation of prediction methods will be obvious if it is remembered that the devices of the type discussed are developed and validated with respect to specific criteria, using available data, in a specific jurisdiction, during a specific time period. Thus, any generalizations to other outcomes of interest, or after modifications of the item definitions used, or to other jurisdictions or populations, or to other time periods, are to be questioned.

3 Which Actuarial Approach?

The general aim of this study was to assess the advantages and disadvantages of some of the most commonly used (or promising) methods for developing statistical prediction methods. Thus, we may ask which of the several available methods "works best." By "best," we refer to predictive efficiency in a validation sample, and, for this assessment, we may compare obtained mean cost ratings. Additional criteria can be used, however; for example, if two procedures give equally valid results, then we may prefer the simpler or less costly method.

Both Meehl[26] and Gough[27] have given good reviews of specific actuarial methods that have been used widely in the behavioral sciences in general, often with particular reference to problems of criminal justice. Mannheim and Wilkins[28] and Simon[29] have provided reviews of specific methods in criminology. The latter author includes various comparisons of predictive efficiency resulting from use of different methods for combining predictors. The present study compared six such methods: two general linear additive models, three clustering models, and a recent multidimensional contingency table (log-linear) approach. A brief and general consideration of each model follows.

Linear Additive Models

Perhaps the most widely used actuarial predictive method has been the linear additive model. Indeed, De Soto has demonstrated that social scientists show a marked proclivity for simple, single linear orderings.[30]

Least-Squares (Multiple) Regression

The best known and most widely used linear additive model is that of least-squares regression, known in the bivariate case as correlation and, in the multivariate case, as multiple correlation or multiple regession. In general, the solution sought is that weighted linear combination of predictor variables that minimizes (the square of) errors about regression to some criterion variable.*

In its general form, the model is described as:

$$\hat{y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_iX_i + \dots + b_nX_n,$$

where (in this case) $\hat{y}$ is the predicted criterion, $\underline{a}$ is some constant (actually the intercept of the regression line), X_i is a predictor variable, and b_i is its weight.

Thus, a potential advantage of the least-squares technique is that it enables estimates of the effect of each predictor variable in terms of a relatively unique (i.e., non-overlapping) contribution to explaining variability in the outcome criterion. This typically results in a set of unequal weights, thus providing an indication (when standardized scores are used) of the relative "importance" of each predictor in the context of all the items used.

*With a dichotomous criterion such as was used in this study, Fisher's discriminant function provides an equivalent method and, therefore, was not included.[31]

While interaction effects (i.e., the extent to which the effect of one predictor variable is mediated or exaggerated by the state of one or several other predictors) can be estimated within a regression framework, the model can quickly become unwieldy.[32]

The power of least-squares regression is beyond question in many applications. As previously discussed, it has been known for some time that least-squares models improve substantially over intuitive clinical approaches to prediction. Moreover, Dawes and Corrigan recently showed that even linear regression models that use *random* regression weights do substantially better than do humans in predictive situations.[33] More recently still, Wainer demonstrated that a simple unweighted (or, more accurately, equally weighted, i.e., all regression weights equal 0.5) linear additive model is essentially as good as, and in some important respects may be better than, a weighted least-squares model.[34]

Studies of parole prediction using data on federal offenders showed that a simple, unweighted model had less shrinkage when applied to validation samples than did the model derived from multiple regression. A variety of possible reasons, some of which relate to the reliability issue, have been discussed by Wilkins.[35] Simon, from her results, came to a similar conclusion.[36] Less shrinkage might be expected from the equal weight model than from a weighted model, simply as a consequence of item unreliability. The weighted models tend to rely heavily on a few items, so if one is scored incorrectly, there is a large difference in the prediction.

Other reasons for differential expected shrinkage are apparent. First, the weighted model may tend to *overfit* the original construction sample data to a greater extent than the unweighted model. This is known as capitalization on chance variation.* Second, the weighted model capitalizes on the presence of any data points that deviate markedly from bi- or multivariate normality (outliers). The equal weights method may be meliorative since regression weights are not estimated from the construction sample data.

The Burgess Method

In the criminal justice field, the unweighted additive model used has been patterned after the work of Burgess.[37] In brief, the procedure involves the use of attribute data (or the dichotomization of predictor variables). Resulting attributes then are used in an unweighted, linear additive fashion to predict the criterion classifications.

Thus, the model is specified as:

$$\hat{y} = (a) + X_1 + X_2 + X_3 + \ldots + X_n,$$

where $\hat{y}$ is as before, <u>a</u> is an arbitrary constant, and X_i is some predictor attribute.†

*Note also that both the weights and the data items will have some error, so the error is multiplied when the weight is applied, i.e., $(\beta + e_\beta)(X + e_X)$.

†Although the Burgess technique is essentially a "linear" model, it is difficult to say precisely what is being treated as a straight line.

Thus, the Burgess technique — one strength of which is found in its simplicity — gives equal weight to all predictors even though there may be markedly unequal levels of association between the criterion and the various predictors used. It allows, therefore, no compensation for "overlapping" effects of the predictors, resulting in the further disadvantage that the technique does not give any indication as to which variables are essentially redundant.

Clustering Models

Lack of power and/or shrinkage on validation in regression predictions may be due in part to discrepancies between data characteristics and analytic assumptions. Usually, for instance, regression analyses do not include interaction terms in prediction equations. (Indeed, in a study with a large number of variables, the examination of individual interactions is often impractical without clear theoretical guidelines.) Another possible limitation of regression analysis derives from the calculation of regression coefficients from the matrix of zero-order correlations among variables within the entire sample. The assumption is made, if implicitly, that the indicated relations hold in population subgroups as well, i.e., that the population is in fact homogeneous. (This may be demonstrably false when the correlation matrices for subgroups are examined.)

The use of clustering methods represents, to some extent, an effort to compensate for the limitations of a regression-based model "in circumstances where interactions and heterogeneities might be expected to reduce the power of multiple regression methods."[38] Clustering methods allow for unspecified interactions and heterogeneities that may be present in a population, and can therefore be characterized as *nonlinear*. Each of the configural methods used here proceeds by a process of hierarchical subdivision.[39]

Each technique used differs from the others not in its general concept, but in terms of the specific algorithms used to successively partition the sample into subgroups (the subgroups, of course, are the issue of interest). Thus, within the same sample, different techniques are likely to result in different groupings. While in many cases the differences will not be substantial, if use of different algorithms results in different partitionings early in the process (i.e., with the first, second, or third attribute entered), terminal subgroups probably will be considerably different.

A large number of clustering algorithms is available.[40] Hierarchical clustering schemes are of two types: *divisive* methods proceed by successively partitioning or subdividing the sample into increasingly homogeneous groups, and *agglomerative* methods reverse this process. (The latter start with the individual and successively group or cluster.) Further differences among methods lie in the specific rules for division or clustering, for termination of the process, or for item inclusion.[41]

Predictive Attribute Analysis and Association Analysis are the two methods used in the present study. Both classify individuals on the basis of the possession or lack of specified attributes, thus providing a subgroup identification. One potential advantage of clustering methods lies in their relative lack of restrictive data assumptions. An additional practical advantage may be that "the method of combining information for any category in an expectancy table is much more readily evident to a non-mathematician in the configural tables than in those involving regression scores."[42]

Both Predictive Attribute Analysis and Association Analysis proceed by classifying a heterogeneous population into relatively homogeneous subgroups, thereby minimizing individual variation *within* subgroups, while maximizing variation *between* subgroups. This is the specific objective of Association Analysis; with Predictive Attribute Analysis, however, within-subgroup variation may be expected to be reduced (if not minimized) and between-subgroup variation may be expected to increase.

As Cormack points out, however, "Often the act of classification has a primary purpose. If so, that purpose should be taken into account."[43] Accordingly, Predictive Attribute Analysis and Association Analysis can be distinguished with respect to the criteria by which the sample is successively partitioned. Predictive Attribute Analysis goes directly to the purpose: the aim is to maximize predictive efficiency by classifying individuals in terms of those predictive attributes that are most strongly associated with the criterion classification (e.g., release outcome). Association Analysis, on the other hand, classifies individuals by those attributes that most effectively summarize shared variance on those same attributes *without* respect to any particular criterion (such as release outcome).[44]

Although with Predictive Attribute Analysis one seeks to maximize predictive power in the construction sample, there are two potential risks inherent in the process of subdivision that may capitalize on random variation (sampling error). The first occurs with the selection of the predictor item with the largest associative value (generally chi-squared or the phi coefficients) with the criterion. It is on the basis of this selection that the categorization is made. Since this attribute is selected without reference to the significance of relations *among* associations, the attribute selected may result from sampling error rather than from its strength as a predictor. Second, the number of hypotheses tested at each subdivision (i.e., the number of attributes under consideration) raises the problem of the possible rejection of the null hypothesis when the attribute and criterion are unrelated.[45]

A possible result, then, is that although Predictive Attribute Analysis provides a method for potentially more accurate prediction,

> *the attributes on which this prediction is based are not necessarily those which indicate the greatest general differences between the individuals. Thus, prediction may sometimes be less "meaningful" and less widely applicable than the rather less precise prediction obtained from Association Analysis.*[46]

We would expect, then, that Predictive Attribute Analyses, which are particularly sensitive to overfitting of the construction sample, would evince greater shrinkage on validation than Association Analyses.

Although Association Analysis may be expected to be more stable and less susceptible to overfitting than Predictive Attribute Analysis, it is not necessarily intended to provide predictive classifications. Indeed, it will yield a predictive model only to the extent that the selected attributes are themselves predictors; Association Analysis alone does not ensure predictive efficiency. The classifications resulting from Association Analysis are

> *basically descriptive rather than predictive. They (may), however, be used predictively if the contrasting homogeneous groups of individuals isolated by the method could be shown to have significantly different outcome probabilities.*[47]

Multidimensional Contingency Table Analysis

This technique, developed by Goodman[48] and others, requires few assumptions about the nature of the variables under consideration or about the nature of relations among them. The technique is complex, although its underlying rationale is relatively straightforward; rather than utilizing a multiplicative model to account for potential interactions among predictor variables, it uses logarithms of the odds ratio, resulting in an additive model.

The model: a) inherently allows for nominal level measures, b) can estimate different "weights" for different predictors, c) can conveniently be used to estimate interaction terms, d) does not require the assumption of a particular multivariate distribution for significance testing (as does regression analysis), and e) provides a means of estimating an "optimal" model. These advantages, plus the utility of the model for identifying a parsimonious set of predictors, make it worth a close examination.

Specifically, the log-linear model predicts odds (ratio of successes to failures) as a multiplicative function of parameters of the predictor variables, each of which is measured at the nominal level.* The analytical form of the model is:

$$\Omega_{ijk} = \gamma\, \gamma^{Ai}\, \gamma^{Bj}\, \gamma^{Ck}\, \gamma^{ABij}\, \gamma^{ACik}\, \gamma^{BCjk}\, \gamma^{ABCijk}$$

where A_i, B_j, and C_k denote the state (category) of the predictor variables, A, B, C, respectively; and Ω_{ijk} denotes the odds that obtain when variables A, B, and C are in states i, j, and k, respectively. The first term on the right side of the equation γ is a constant, which represents a basic odds rate (similar to a base rate probability) from which the effects of the predictors are deviations. The second through fourth terms γ^{Ai}, γ^{Bj}, and γ^{Ck}, represent the direct "main effects" of the predictors A, B, and C. The fifth through seventh terms represent two-way interaction effects, and the last term, the three-way interaction effect of the predictors. It should be noted that this equation represents a "saturated" model, i.e., it specifies all possible main effects and interaction effects of the independent variables. In the case of three independent variables, the saturated model will have eight terms (including the constant) on the right side of the equation.

*Other models, which may use different measurement levels, also exist.

Since technically we refer to a model of the *log* of the odds ratio, the equation given below expresses the model in terms of logs:

$$O_{ijk} = T + T^{Ai} + T^{Bj} + T^{Ck} + T^{ABij} + T^{ACik} + T^{BCjk} + T^{ABCijk}$$

where O_{ijk} is the log of Ω_{ijk}, and the T's are logs of the corresponding Υ's.

4 A Comparison of Methods

The specific predictive methods to be compared in the present study are summarized in Table 1 with respect to the several issues discussed in the last section. Specifically considered are: a) the extent to which a given method accounts for predictor variable intercorrelations, b) whether the method assumes linearity and/or additivity of relations, and c) the expected tendency of the model to overfit the construction sample data (with concomitant increase in shrinkage on validation).

The methods selected for use in this comparative study were intended to provide a range of variation in the characteristics discussed. As can be seen from Table 1, each characteristic is present in at least one method and absent in at least one, and the expected tendency for construction sample overfitting varies from low to high.

Table 1
CHARACTERISTICS OF SIX PREDICTION METHODS

Method	Account for Predictor Intercorrelation?	Linear Relations Assumed?	Additive Relations Assumed?	Expected Tendency for Overfitting
Burgess	No	Yes[a]	Yes	Low
Multiple Regression	Yes	Yes	Yes	Moderate/High
Association Analysis	Yes	No[a]	No	Low
Association Analysis with Criterion-Referenced Decision Rules	Yes	No[a]	No	Moderate
Predictive Attribute Analysis	Yes	No[a]	No	High
Multidimensional Contingency Analysis	Yes	No	Yes[b]	Moderate/High

Notes: [a]Since all variables are dichotomized, the issue of linearity is essentially ignored. Assumptions of linearity are needed given the Burgess technique; they are not given the hierarchical clustering techniques.

[b]The model is actually multiplicative; it is, however, additive in the log odds ratio.

[c]A version of this table first appeared in an unpublished report prepared by K. Andreason, W. Brown, G. Dodsley, M. Neithercutt, G. Pasella, D. Pfoutz, and S. Springer of the National Council on Crime and Delinquency Research Center.

The Data Base

A major cost of this type of study is the collection and preparation for analysis of appropriate data. The present study avoided this problem by making use of a large data base initiated under the Parole Decisionmaking Project.[49]

The file used contains data on more than 4,500 people released from federal prisons in the years 1970 through 1972. Of these, approximately 2,400 were released in calendar year 1970; 1,000, in calendar year 1971; and 1,100, in calendar year 1972. The sample does not reflect all releasees, and different proportions were sampled in different years: the 1970 set is a 50 percent sample of persons released during the first six months and a 20 percent sample of those released during the last six months of the year; the 1971 set is a 30 percent sample of persons released during the last six months of the year; and the 1972 set is a 30 percent sample of those released during the first six months of the year. All samples were drawn in a manner such that the observations may be assumed to approximate those that would be obtained via random selective procedures.

At least two years of follow-up data are available for all three samples, and systematic "track-down" was accomplished for releasees for whom no disposition was indicated on FBI arrest records.

More than 90 items of information concerning each offender were recorded during the Parole Decisionmaking Project. These included social and criminal history; the nature and circumstance of the present offense (i.e., the offense that resulted in the period of confinement from which the subject was released during the specified project year) and incarceration; institutional adjustment and custody classifications; and the nature of the offender's post-release adjustment.

Information was coded from case files, parole reports, or FBI arrest records. The data base was refined from that available from the parole commission in order to increase its utility for a variety of research purposes.

Construction and Validation Samples

Given our use of this data base, the only major issue of sampling with which we had to be concerned involved splitting the sample into useful construction and validation sub-samples. In view of differences in data collection procedures and in the availability of information across years, it was decided that the 1970-release sample would serve as the construction sub-sample and that the 1972-release sample would serve as the validation sub-sample.

After a series of preliminary analyses, it was decided also that the entire 1970 and 1972 samples would be used (including any females and/or juveniles released). The analyses which led to this decision are described subsequently.

The Criterion Problem

Post-institutional adjustment is a highly complex concept. In examining it, one could include such variables as employment status, familial relations, or personal satisfaction, as well as measures of recidivism. Moreover, recidivism is itself a complex variable. How does one define behavior which is "recidivistic" as opposed to that which is not? In essence, we are faced with a classification problem. The process of classification is not simple, although it is often treated in a relatively simple manner.

Classification is that procedure which forms the basis for all measurement; classification and measurement are different processes. We cannot measure objects or events; rather, we can only measure attributes or properties of objects or events. Thus, we cannot measure a phenomenon such as burglary. We can measure the

seriousness of burglary, or its frequency, or the dollar cost associated with it — the attributes or properties of burglary — but not the phenomenon itself.

While we cannot measure burglary other than through its attributes, we can classify burglaries: commercial vs. residential, for example. Indeed, classification must precede measurement in a logical sequence. In order to measure some property of the class of phenomena known as burglary, we must first determine whether any given event is in fact a burglary.

Classification can be defined as that process whereby objects, events or phenomena become *categorized* (typically for the purpose of counting). Only two simple, but fundamental, rules must be followed in this process: 1) the categories into which we classify must be mutually exclusive (i.e., an event can be placed in one and only one category within the classification scheme), and 2) the set of categories developed must be exhaustive (i.e., all occurrences of the event must be accommodated by the classification scheme). Since these are the only two rules required, anytime one can make a legitimate distinction among events, one can create a new — and perfectly legitimate — classification scheme. A major problem facing those who would develop parole risk-screening devices, then, is the classification of outcome variables.

In the Classification for Parole Decisionmaking (PDM) Project,[50] the outcome criterion had been defined as follows. If, within the follow-up period, there was a) a return to prison with either a new offense conviction or a technical violation, or b) an outstanding warrant for absconding from supervision, or c) a sentence to confinement for a period of 60 days or more, or d) the subject had died as a result of a criminal act, then an offender was classified as a "failure." All other cases were classified as "successes."

In practice, three separate outcome variables must be examined in order to arrive at this classification when using the PDM data base. Table 2 describes these variables and their category definitions. Clearly, other potential criteria could be developed given these three outcome-related variables.

Four criterion variables are developed in Table 2. Criterion A is that utilized in the PDM studies.* Essentially, the PDM scheme searches for the specific categories of misconduct noted above. If none is found, then "success" is presumed.

The remaining three outcome criteria differ from the PDM classification in one or more ways. Criterion B, for example, counts *any* known conviction as a failure (as opposed to only convictions resulting in a sentence of 60 days or more of confinement). We should note here that the data set used results in some unknown degree of bias given this criterion, since one critical outcome variable (parole performance) includes the category "Continued on Parole (no difficulty or sentence less than 60 days)" and does not allow us to distinguish between those subjects who encountered no difficulty and those who, in fact, may have had one or more convictions resulting

*It should be noted that, with modification of the recoding scheme, a somewhat different distribution of the criterion could easily occur. Substantive conclusions, however, should remain the same.

in sentences of less than 60 days. Further, subsequent outcome measures, such as "Number of Convictions," were not coded for these cases. While the measure is thus biased, and the extent of the bias is unknown, it can also be argued that the original PDM criterion (our Criterion A) is itself biased against "lesser failures." If this is the case, then Criterion B is potentially somewhat less subject to this bias.

Criterion C is essentially the same as Criterion B, with two exceptions. First, subjects for whom no clear outcome, either successful or otherwise, can be identified are omitted from the sample. Second, subjects known to have been arrested for, but not necessarily convicted of, an offense are considered as failures. With respect to this criterion, we should note that an additional potential bias is introduced, since arrest information is available usually only for those subjects for whom FBI follow-up information is available.

Finally, Criterion D duplicates Criterion C, with the exception that we return to the conviction criterion (i.e., "arrests only" are now considered as successful).

For purposes of preliminary analysis, we utilized Criterion A — that decided upon for the original PDM studies — although any would be acceptable for use, depending upon one's purpose in instrument development.* It should be noted that the sample is not limited to persons paroled, but includes also those conditionally released and discharged. There is, therefore, a differential exposure to the risk of being counted as a "failure" associated with the mode of release. Persons conditionally released may be expected to serve, on the average, less time under supervision, while those discharged will not be supervised at all. Those under supervision are subject to return to prison for technical violations, but those discharged before the end of the follow-up period are not subject to this classification after that.

Given this initial criterion, the base rate for successful follow-up performance for the 2,467-case 1970 sample is 71.5 percent. For those 1,159 subjects who were released under parole supervision, the "success" rate is 75.2 percent. For those 717 subjects who were mandatorily released under supervision, the "success" rate is 65.6 percent. Finally, for the 591 subjects who were released at the expiration of their sentence and without supervision, the "success" rate is 71.5 percent. (Note that, in view of the bias in classification just discussed, and in the circumstance of potential parole selection effects, these differences in rate are merely descriptive and do not comprise a comment on the relative "effectiveness" of different modes of release.)

*This decision was made for two reasons: 1) it allowed us to replicate analyses performed by others, thereby validating our revisions of the data base, and 2) it had originally been agreed upon as reasonable by members of the parole comission—an important consideration in the development of statistical risk-screening devices designed for operational use. As described in a later section, however, Criterion D was chosen for final instrument construction and validation.

Table 2
OUTCOME CODINGS AND DEFINITIONS

Outcome-Related Variables	Potential Criterion Codings: Criterion A	Criterion B	Criterion C	Criterion D
DEATH (First Pass Variable)				
0 Subject was alive or presumed alive at the end of the follow-up period (N = 2,471)	Examine FBI Outcome Variable			
1 Subject died or is presumed to have died before the end of the follow-up period while on parole (no criminal act was involved) (N = 17)	F	F	—	—
2 Subject died or is presumed to have died after release from parole but during the follow-up period (no criminal act was involved) (N = 2)	F	F	—	—
3 Subject died or is presumed to have died before the end of the follow-up period while on parole (a criminal act was involved) (N = 7)	U	U	U	U
4 Subject died or is presumed to have died after release from parole, but during the follow-up period (a criminal act was involved) (N = 0)	U	U	U	U
FBI OUTCOME (Second Pass Variable)				
0 No entry during follow-up period (N = 1,436)	Examine Parole Performance Variable			
1 FBI record is not located (N = 1)				
2 Parole or mandatory release violation—return to prison, no new conviction (N = 163)	U	U	U	U
3 Conviction resulting in sentence of 60 days or more is indicated (N = 493)	U	U	U	U
4 Conviction resulting in sentence of less than 60 days is indicated (N = 122)	F	U	U	U
5 Conviction is noted, but sentence is not specified (N = 3)	F	U	U	U
6 Arrest(s) indicated, but no convictions or returns to prison noted (N = 146)	F	F	U	F
7 Parole or mandatory release violation—no return to prison (N = 15)	F	F	U	F
8 Unknown and unspecified code (N = 86)	F	F	—	—
9 Other (N = 6)	F	F	—	—
PAROLE PERFORMANCE (Third Pass Variable)				
0 **Continued on Parole (no difficulty or sentences less than 60 days)** (N = 1,068) Subject has not absconded from parole, has no minor or major convictions, and no actions as described in the following codes have been taken by the parole authority. Note that the subject may have had one or more convictions resulting in sentences of less than sixty days confinement each, with or without actual confinement, suspended sentence, or probation.	F	F	F	F

Note: U = Unfavorable
F = Favorable

Table 2 (continued)
OUTCOME CODINGS AND DEFINITIONS

	Potential Criterion Codings			
	Criterion A	Criterion B	Criterion C	Criterion D
PAROLE PERFORMANCE (continued)				
1 **Continued on Parole (new minor conviction(s))** (N = 18) Subject has been continued on parole after one or more minor convictions, for one or more offenses, committed while on parole. Note that a minor conviction means that the subject received a sentence of at least sixty days but less than one year, whether or not the sentence resulted in actual confinement, suspended sentence, or probation.	F	U	U	U
2 **Absconder** (N = 13) The whereabouts of the parolee are unknown to the paroling authority. Either a warrant for absconding from parole has been issued or some other official action has been taken to declare the parolee an absconder. If by policy no official acts are customarily taken with respect to absconders, then this code should be used when the parolee has been out of contact more than two months and his or her whereabouts are clearly unknown.	U	U	U	U
3 **Returned to Prison—Technical Violation (no new conviction(s) and not in lieu of prosecution)** (N = 4) The parolee has been declared a parole violator by the paroling authority and returned to prison. No criminal convictions (major, minor, or lesser) occurred during parole. This code includes those who are returned: Simply for absconding from parole. For failure to follow other parole rules. For further treatment (including psychiatric but excluding medical) related to their parole performance. Under treatment and control programs, such as narcotic users, alcoholics, or any others who are adjudged to need further institutional treatment before discharge or continuance on parole.	U	U	U	U
4 **Returned to Prison—Technical Violation (new minor or lesser conviction(s) or in lieu of prosecution on new minor or lesser offense(s))** (N = 4) The paroling authority has declared the parolee to be a parole violator **and** the parolee has committed an offense for which the maximum sentence is less than one year. The parolee has been returned to prison **either** after having been convicted and sentenced, including suspended sentence or probation, **or** in lieu of prosecution and on the basis of a clear admission of guilt for an offense which, if successfully prosecuted, would have resulted in a maximum sentence of less than one year.	U	U	U	U

Table 2 (continued)
OUTCOME CODINGS AND DEFINITIONS

	Potential Criterion Codings			
	Criterion A	Criterion B	Criterion C	Criterion D
PAROLE PERFORMANCE (continued)				
5 **Returned to Prison—Technical Violation (in lieu of prosecution on new major offense(s))** (N = 5) The paroling authority has declared the parolee to be a parole violator **and** the parolee has committed an offense for which the maximum sentence is at least one year. The subject has been returned to prison in lieu of prosecution and on the basis of a clear admission of guilt for an offense which, if successfully prosecuted, would have resulted in a maximum sentence of one year or more. Whenever this code is used, the appropriate code for this "New Offense" should be entered in columns 30-31 below.	U	U	U	U
6 **Returned to Prison—No Violation** (N = 0) The subject has been returned to prison for reasons not reflecting on his or her performance since paroled. Examples are: Return for medical reasons other than psychiatric. Return on a new commitment for an offense committed before release on parole.	U	U	U	U
7 **Recommitted to Prison—New Major Conviction(s) (same jurisdiction)** (N = 1) The subject has been convicted, sentenced, and recommitted to prison, or given a suspended sentence or probation, in the same jurisdiction for offense committed since he or she was paroled, with a maximum sentence of at least one year.	U	U	U	U
8 **Recommitted to Prison—New Major Conviction(s) (any other jurisdiction)** (N = 19) The subject has been convicted, sentenced and committed to prison in any other jurisdiction, i.e., to state, territorial or foreign prisons. The offense(s), committed since subject was paroled, has a maximum sentence of at least one year.	U	U	U	U
FINAL PASS: Ensure that any remaining cases have a code of 0 for FBI outcome FBI code 0 (N = 301)	F	F	F	F
FBI code 1 or blank (N = 4)	—	—	—	—

The Data Reduction Problem

As a first step toward the development of any statistical risk-screening device, one must somehow delimit the range of variables that could be used in its construction. Clearly, this process actually begins long before the stage of data analysis — decisions are made about what data (which variables) to collect and in what form. At this stage of preliminary analysis, then, one seeks to further reduce the size of the data set without sacrificing important information.

There are, of course, many ways in which one could go about such a task. For example, if we were interested only in data reduction, a factor analytic approach might be appropriate. Since our study has a clear, single purpose, however — to predict post-release criminal behavior — some method that allows reduction, but that does so while attending to relations with the criterion variable, was desired.

Accordingly, our first step in this process was to compare, for each item, the joint (bivariate) distributions of each potential predictor variable with the outcome. In addition to providing critical base-line information about the sample, such a careful item-by-item examination provides the researcher with a wealth of information that can be of use in the development of statistical screening devices.

The first benefit is that this procedure can serve as a safeguard against the unwitting violation of certain statistical assumptions necessary for the appropriate use of some predictive methods. For example the multiple-regression method requires multivariate normality and homogeneity of variance, and its use could result in a misleading analysis if extreme scores (outliers) were not appropriately handled.

Further, the various statistical methods used in this study require different characteristics of predictor variables. Some, like the Burgess technique, demand (typically) that variables consist of dichotomies. Others, like multiple regression, require that variables possess characteristics of interval scales. Still others, like the log-linear techniques, allow for the inclusion of multi-category, nominal-level variables. Thus, the majority of the items have been transformed one or more times with the requirements of the various techniques in mind. For example, while the log-linear methods allow for the use of multi-category, nominal-level data, very large numbers of categories are problematic for two reasons: 1) the size, in terms of numbers of cells, of the n-dimensional table can quickly become too large to handle, and 2) zero frequency cells are problematic if this method is used. Hence, a balance between too many categories and too few categories (to allow for the adequate consideration of interaction) must be sought.

A final advantage of the studies is that they allow one to examine, in a relatively rough fashion, the effect of subgroup heterogeneities. For example, it may well be the case that certain variables might be differentially predictive for men, women, or juveniles. These specific subgroups all were examined in this fashion and results are given in Appendix 3. No major differences were noted; hence, the remainder of the studies reported in this paper deal with the full sample of releasees.

Based on these analyses, the original data set was reduced to a subset of 30 items that appear to have both predictive and practical utility for the development of risk-screening devices. Table 3 lists this subset of variables and the transformations (if any) used.*

Most of the variables under consideration fall into one of four general categories: 1) variables relating to the present offense, 2) variables relating to a subject's history

*One variable (dollar amount associated with the present offense) was subsequently discovered to have been excluded from the 1972 data set. Hence, it appears in no analyses subsequent to those reported in this chapter.

of criminal or delinquent involvement, 3) "social history" variables, such as the highest completed grade claimed, or living arrangements prior to commitment, and 4) variables relating to institutional adjustment, during either the present or some past incarceration. (One selected variable, "Planned Living Arrangement," while used in some analyses, does not easily fall into any of these four categories.)

Table 3
VARIABLE DESCRIPTIONS, BY GROUP

Variables	Coding
RELATING TO PRESENT OFFENSE	
How Committed	Probation, Parole, or Mandatory Release Violation vs. All Other Commitments
Type of Admission	New Court Commitment vs. All Other Types of Admissions
Commitment Offense	a) Burglary, Larceny, Theft (Any Type) vs. All Other Offenses b) Homicide, Manslaughter, All Robbery, All Theft, Larceny, Fraud, Forgery, Counterfeiting, Kidnapping vs. All Other Offenses (Person and Property vs. Other) c) Vehicle Theft, Forgery, Fraud, Larceny by Check, Heroin vs. All Other Offenses d) Vehicle Theft, Forgery, Fraud, Larceny by Check vs. All Other Offenses
Dollar Value of Offense	a) Dollar Value b) Unknown or Less than or Equal to $499 vs. $500 or More
RELATING TO CRIMINAL HISTORY	
Age at First Arrest	a) ≤ 14; 15-17; 18-21; ≥ 22 b) ≤ 18; > 18
Age at First Conviction	a) ≤ 15; 16-18; 19-22; ≥ 23 b) ≤ 18; > 18
Age at First Commitment	a) ≤ 17; 18-20; 21-25; ≥ 26 b) ≤ 18; > 18
Longest Time Free Since First Commitment	a) ≤ 6 mos; 7-18 mos; 19-36 mos; 37-60 mos; > 60 mos (includes code of '0') b) Code '0' and > 60 mos vs. ≤ 60 mos
Number of Prior Convictions	a) None; 1; 2-3; 4 or More b) None vs. Any
Number of Prior Sentences	a) None; 1-2; 3 or More b) None vs. Any
Number of Prior Sentences with Probation	a) None, '1'; 2 or More b) None vs. Any
Number of Prior Incarcerations	a) None; '1'; 2 or More b) None vs. Any
Number of Prior Parole/Probation Revocations	None vs. Any

Table 3 (continued)
VARIABLE DESCRIPTIONS, BY GROUP

Variables	Coding
RELATING TO CRIMINAL HISTORY (continued)	
Number of Prior Convictions for Burglary	a) Number of Prior Convictions for Burglary b) None vs. Any
Number of Prior Convictions for Larceny	a) Number of Prior Convictions for Larceny b) None vs. Any
Number of Prior Convictions for Auto Theft	a) Number of Prior Convictions for Auto Theft b) None vs. Any
Number of Prior Convictions for Forgery, Fraud, or Larceny by Check	a) Number of Prior Convictions for Forgery, Fraud, or Larceny by Check b) None vs. Any
Total Number of Prior Convictions for Property Offenses	a) 0-1; 2; 3; 4 or More b) 0-1; 2 or More
Longest Time Served on Any Commitment	0-6 mos vs. more than 6 mos
Reason for First Arrest	Burglary, Check Offenses, Forgery, Theft, Delinquent Child vs. All Others
RELATED TO SOCIAL HISTORY	
Highest Grade Completed	a) None through Ph.D. b) 0-11; 12 or Greater
Marital Status at Admission	Married or Common-Law vs. Any Other Status
Living Arrangement Before Commitment	Wife and/or Children vs. Any Other Arrangement
Use of Synthetic and/or Natural Opiates	Known Use vs. No Known Use
Employment During Last Two Years of Civilian Life	Employed more than 25% of the time; or Student; or Unemployable 75% of the time vs. Unemployed
Longest Job in Free Community	a) Unknown or less than one year; 1-4 years; more than 4 years b) Unknown or 4 or fewer years vs. more than 4 years

Table 3 (continued)
VARIABLE DESCRIPTIONS, BY GROUP

Variables	Coding
RELATED TO INSTITUTIONAL ADJUSTMENT	
Escape History	No Escapes or Attempted Escapes from Any Custody vs. One or More Escapes or Attempted Escapes
Latest Custody Classification	a) Maximum or Close; Medium; Minimum; Work-Release or Unknown b) Unknown, Minimum, or Work-Release vs. Maximum, Close, or Medium
Prison Punishment	None vs. Any
ADDITIONAL VARIABLE	
Planned Living Arrangement	Wife and/or Children vs. Any Other Arrangement

The Problem of Multicollinearity

Some of the analyses to be performed require that predictor variables themselves be relatively orthogonal. In particular, highly intercorrelated predictor variables (or subsets of predictor variables) can result in multiple-regression equations that may be misleading for practical prediction. Tables A-1 through A-7 give, by variable category, the intercorrelation matrices of the 29 selected variables and, in some cases, their transformations. (The "Planned Living Arrangement" item was excluded.) Inspection of these tables demonstrates that multicollinearity does not present a substantial problem for this data set, although examination of these matrices resulted in the deletion of a few highly correlated variables from the regression analyses.

The different analyses to be compared require different characteristics of predictor variables. For simplicity, we treated all ordered, multi-category variables as "continuous" and others were dichotomized. Although the treatment of some of these ordered, multi-category variables as continuous violates some statistical assumptions (e.g., in the calculation of correlation coefficients), the particular statistical techniques we used are remarkably robust. This is especially true with respect to the particular assumptions violated (e.g., the assumption of equality of intervals), provided the violation is not unreasonable.

Tables A-8 through A-14 summarize relations between the selected variables or attributes and the four outcome criteria outlined in Table 2. By and large, relations of individual items with the outcome criteria are low (although, with this large number of observations, all are statistically significant). Overall, Criteria B, C, and D seem a bit more related to the predictor variables chosen for further study, even though all variables were selected with reference to their correlation with Criterion A. Of the four, Criterion D appears best related to all variables, although its advantage over its competitors is not significant. Given the slight empirical advantage, however, and its intuitive appeal, Criterion D was used for the analyses that follow.

Linear Additive Models

Burgess

As a first step toward the development of a Burgess scale for the assessment of risk, all 29 dichotomous variables that had been found to be both significantly (in a statistical sense) and substantially (in a practical sense, relative to other items) associated with the criterion were included. Table 4 gives the joint distribution of outcome and Burgess scale scores for the instrument developed using all 29 items. The point-biserial correlation coefficient between scale scores and the outcome measure is 0.35; the Mean Cost Rating is 0.43.* Both are highly significant statistically.

For such a device to be operationally useful, it is often advisable to collapse the Burgess scale score distribution into a smaller number of categories. This serves several functions. First, Table 4 clearly shows that, in general, higher Burgess scores are associated with a greater probability of favorable outcome. There are, however, a number of reversals. For example, it might appear from the table that an individual with a Burgess score of 5 would have a 50/50 chance of recidivating. Higher values of the Burgess scale are supposed to reflect increasing likelihood of a successful parole outcome; yet, if we move up to a score of 6, we find that the associated probability of favorable outcome is only 0.293! In fact, we have to move quite a way up the scale before we again increase the likelihood of success over 0.5. The same thing can and, in this case, does happen at the higher values of the scale.

While this sort of occurrence poses no statistical problem, it can, of course, result in important operational problems. Collapsing the score distribution can, and often will, ameliorate these.

Collapsing the score distribution serves a second function related to the first. Typically, these reversals can be ascribed to random variation within the construction sample. That is, were we to construct a similar instrument using a different sample of offenders, we would again expect to observe some reversals, but they typically would not fall at the same point(s) on the scale. In collapsing score categories, we are assuming that the "bumps and wiggles" in the distribution are not "real" or "true" — and, hence, they are unimportant.

Collapsing the distribution can be expected to have two related effects. First, the correlation coefficient (or other measure) used to assess the relation between the scale and the criterion can be expected to decrease in magnitude. Second, however, we should expect to see less shrinkage of this coefficient on cross-validation. Thus, while it might appear that our predictive power is lessened through the collapsing procedure, actual predictive power (while it may also decrease somewhat) should be less affected.

*For reasons discussed in an earlier section, the MCR, which is based on a cutting score approach and takes account of the base rate of success (or failure), was the statistic selected for use in comparing different methods.

Having determined that it is desirable to collapse the score distribution, we are faced with decisions as to how to go about it. A number of schemes are potentially appropriate. We could, for example, simply collapse adjacent categories such that the transformed categories are each constructed of the same number of original score categories. We shall call this the "equal step" method. It is illustrated in Table 5. As expected, both r_{pb} and MCR decrease a bit.

An alternative approach is illustrated in Table 6. Here, rather than equal-appearing scale intervals, we sought to ensure roughly equal numbers of subjects in the transformed categories. In this example, we have used pentiles (20th, 40th, 60th, 80th, and 100th percentile point boundaries) to define the category boundaries, but any other such scheme (e.g., based on deciles) could serve as well.

A third, and potentially more justifiable, approach is illustrated in Table 7. Although the original distribution of the Burgess scale scores is positively skewed, the skew is not substantial, and, in fact, the distribution provides a rough approximation to the normal. Hence, we have in Table 7 divided the original distribution (roughly) into half-standard deviation units.

As can be seen from Tables 5 through 7, all collapsing schemes have had both the desired and expected results with respect to reversals and power. How then to choose among them? In practice, and as can be seen from the coefficients, any choice would make little difference. There is still the validation issue to be considered, but, again, little practical difference in coefficients is likely to be observed. One might well suspect, however, that collapsing schemes that make use of properties of the distribution itself (such as the pentile or standard deviation schemes) rather than the proclivity of the developer (such as the "equal step" method) might be more stable.

Another issue of concern in the development of risk-screening devices designed for operational use involves the number of items from which the device is constructed. Again, we are faced with a trade-off situation. As the number of items used increases, so, in general, does the reliability of the assessment device. As the number of items increases, however, so does the effort involved in data gathering and assessment.

Table 8 summarizes a Burgess-type device developed from the 19 variables (out of the original 29) for which the first-order correlation coefficients with the outcome measure met or exceeded |0.15|.* Tables A-15 through A-17 summarize the results of collapsing strategies analogous to those discussed above.

*The 19 items are: all burglary, larceny, theft and fraud offenses; age at first arrest; age at first conviction; age at first commitment; longest time free since first commitment; number of prior sentences; number of prior incarcerations; number of prior parole/probation revocations; number of prior convictions for burglary; number of prior convictions for auto theft; longest time served on any commitment; marital status at admission; living arrangement before commitment; employment in last two years of civilian life; longest job in free community; escape history; prison punishment; vehicle theft, burglary, larceny, forgery, fraud and counterfeiting; and planned living arrangement.

Validation Study. Each of the Burgess-type devices described in the preceding section was validated on the 1972 data set. Burgess scores of each type were calculated for each of the 1,004 1972 releasees, and these scores were correlated with the outcome criterion.

Table 9 summarizes these findings. Very little shrinkage is apparent. As expected, the Burgess-type devices — in all their forms — hold up very well on validation. Finally, it can be seen that virtually nothing (including stability) is gained by the inclusion of 10 additional items (in the 29-item devices).

Multiple-Regression Analyses

A common finding in multiple-regression research is that little predictive power is provided by the inclusion of more than the first few variables in the equation. That is, the first few variables to enter the equation explain the bulk of the variation that the equation eventually will explain.

Table 10 summarizes the regression equation obtained from 18 variables (of the available 29) that contributed statistically significant increments to R when variables were free to enter in any order. Clearly, the first five or six variables "account for" the bulk of the explained variation in parole outcome, and little useful information is gained from the rest.

Most of the 29 selected variables were classified, as already noted, into four categories relating to the present offense, criminal history, social history, and institutional adjustment. Tables A-18 through A-21 summarize each of the regression equations obtained when using only variables from one such subset.

Using only variables related to the present offense (Table A-18), we are able to account for only 4 percent of the variation in outcome. Note, however, that this equation is based on only two variables. (A third present offense variable, "How Committed," is perfectly correlated with "Type of Admission," and including both would result in substantial problems of collinearity.)

Table A-19 summarizes the equation developed using only variables related to offenders' criminal histories. Here, our predictive ability has increased substantially. Twelve percent of the outcome variance can be accounted for given knowledge of the eight variables included in the equation.

Six percent of the outcome variance is explained using only six variables related to offenders' social histories (Table A-20). Finally, the three variables relating to offenders' institutional adjustment allow us to account for 5 percent of the variation in outcome (Table A-21).

While the equation presented in Table 10 includes variables from all four groups, it was developed without regard to variable group classification. Rather, the entry of a given variable into the equation was dependent only upon its correlation with the criterion and its intercorrelation with the other variables.

The items one may wish to include in a risk-screening device may depend, as discussed in a previous section, not only on their contribution to explained variance

but also on issues of data availability and policy. For example, items reflecting institutional adjustment will not be available at the time of reception in prison, and, depending upon the intended use of the device, issues of fairness may arise with respect to the justifiable use of information other than the offense and perhaps prior criminal record. It is potentially of interest, therefore, to assess the viability of equations developed with regard to the variable classifications used.

Table 4
TWENTY-NINE BURGESS SCALE/OUTCOME DISTRIBUTION

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
3	1	100.0
4	19	21.1
5	38	50.0
6	58	29.3
7	74	45.9
8	133	41.4
9	131	47.3
10	159	52.2
11	160	56.3
12	148	53.4
13	145	53.8
14	158	54.4
15	123	66.7
16	128	70.3
17	124	66.9
18	107	76.6
19	92	75.0
20	76	85.5
21	80	81.3
22	84	78.6
23	66	84.8
24	78	91.0
25	82	93.9
26	31	96.8
27	38	100.0
28	43	97.7
29	6	100.0
TOTAL GROUP	2,382	64.2

Eta = 0.365
Eta^2 = 0.133

r_{pb} = 0.352
r_{pb}^2 = 0.124

MCR = 0.4287

Note: Outcome is defined as follows (see text and Table 2 for a complete description):

Favorable Category

a) Continued on parole (no difficulty or sentences less than 60 days) (N = 1,068)
b) No FBI entry during follow-up period (N = 301)
c) Arrests, but no convictions or returns to prison noted (N = 146)
d) Parole or mandatory release violation, no return to prison (N = 15)

Unfavorable Category

Any return to prison, or
Any conviction for any new offense, or
Death during the commission of a criminal act, or
Absconding from supervision

Table 5
TWENTY-NINE BURGESS SCALE/OUTCOME DISTRIBUTION EQUAL-STEP COLLAPSE

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
3-5	58	41.4
6-8	265	40.0
9-11	450	52.2
12-14	451	53.9
15-17	375	68.0
18-20	275	78.5
21-23	230	81.3
24-26	191	93.2
27-29	87	98.9
TOTAL GROUP	2,382	64.2

Eta = 0.356
Eta^2 = 0.127

r_{pb} = 0.350
r_{pb}^2 = 0.123

MCR = 0.395

Table 6
TWENTY-NINE BURGESS SCALE/OUTCOME DISTRIBUTION PENTILES

(Construction Sample)

Burgess Score (Pentile)	N	Percent Favorable Outcome
3-9	454	42.3
10-12	467	54.0
13-16	554	60.6
17-20	399	74.9
21-29	508	88.8
TOTAL GROUP	2,382	64.2

Eta = 0.338
Eta^2 = 0.114

r_{pb} = 0.336
r_{pb}^2 = 0.113

MCR = 0.414

Table 7
TWENTY-NINE BURGESS SCALE/OUTCOME DISTRIBUTION HALF-STANDARD DEVIATION UNITS

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
3-6	116	35.3
7-9	338	44.7
10-12	467	54.0
13-15	426	57.7
16-18	359	71.0
19-21	248	80.2
22-24	228	84.6
25+	200	96.5
TOTAL GROUP	2,382	64.2

Eta = 0.352
Eta^2 = 0.124

r_{pb} = 0.349
r_{pb}^2 = 0.112

MCR = 0.412

Table 8
NINETEEN-ITEM BURGESS SCALE/OUTCOME DISTRIBUTION

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
0	13	15.4
1	52	30.8
2	99	36.4
3	141	42.6
4	172	55.8
5	173	53.8
6	184	47.8
7	211	61.1
8	182	54.9
9	172	59.3
10	150	69.3
11	152	76.3
12	142	80.3
13	116	81.9
14	111	81.1
15	122	86.9
16	61	95.1
17	47	97.9
18	66	95.5
19	16	100.0
TOTAL GROUP	2,382	64.2

Eta = 0.364
Eta^2 = 0.132

r_{pb} = 0.350
r_{pb}^2 = 0.123

MCR = 0.429

Table 9
BURGESS DEVICE VALIDATION STUDIES

Burgess Device	Point-Biserial Correlation (r_{pb}) Construction Sample (N = 2,382)	Point-Biserial Correlation (r_{pb}) Validation Sample (N = 1,004)	Coefficient of Determination (r_{pb}^2) Construction Sample (N = 2,382)	Coefficient of Determination (r_{pb}^2) Validation Sample (N = 1,004)	Mean Cost Rating (MCR) Construction Sample (N = 2,382)	Mean Cost Rating (MCR) Validation Sample (N = 1,004)	Shrinkage rpb^2	Shrinkage MCR
Twenty-Nine Items								
Uncollapsed Version	0.352	0.335	0.124	0.112	0.429	0.444	0.012	—
Equal-Step Collapse	0.350	0.330	0.123	0.109	0.395	0.407	0.020	—
Pentile Collapse	0.336	0.332	0.113	0.110	0.414	0.394	0.003	0.020
Half-Standard Deviation Collapse	0.349	0.335	0.122	0.112	0.412	0.406	0.010	0.006
Nineteen Items								
Uncollapsed Version	0.350	0.333	0.123	0.111	0.429	0.423	0.012	0.006
Equal-Step Collapse	0.343	0.320	0.118	0.102	0.394	0.371	0.016	0.023
Pentile Collapse	0.332	0.319	0.110	0.102	0.389	0.378	0.008	0.011
Half-Standard Deviation Collapse	0.345	0.335	0.119	0.112	0.408	0.404	0.007	0.004

Table 10
MULTIPLE REGRESSION OF 18 PREDICTOR VARIABLES ON OUTCOME[1]

Variable[2]	Unstandardized Weight	Standardized Weight	R	R^2
Longest Time Free	-0.0307	-0.1049	0.269	0.073
Crime Group	0.0292	0.0710	0.318	0.101
Prison Punishment	0.0957	0.0902	0.336	0.113
Living Arrangement	0.0578	0.0507	0.349	0.122
Age at First Arrest	-0.0613	-0.0632	0.356	0.127
Known Use of Synthetic and/or Natural Opiates	0.0548	0.0852	0.362	0.131
Type of Admission	0.0401	0.0352	0.367	0.135
Number of Prior Commitments for Auto Theft	0.0257	0.0648	0.370	0.137
Employment	0.0349	0.0364	0.373	0.139
Escape	0.0411	0.0344	0.375	0.140
Highest Grade Claimed	-0.0064	-0.0375	0.376	0.142
Commitment Offense	-0.0469	-0.0478	0.378	0.143
Reason for First Arrest	0.0442	0.0461	0.380	0.144
Planned Living Arrangement	0.0524	0.0434	0.381	0.145
Number of Prior Commitments for Forgery	0.0198	0.0437	0.382	0.146
Custody Classification	0.0183	0.0295	0.384	0.147
Number of Prior Commitments for Burglary	0.0299	0.0281	0.384	0.148
Number of Prior Parole/Probation Revocations	0.0155	0.0317	0.385	0.148
Constant	1.2646			

Outcome is defined as follows (see text and Table 2 for a complete description):

Favorable Category

a) Continued on parole (no difficulty or sentences less than 60 days) (N = 1,068)
b) No FBI entry during follow-up period (N = 301)
c) Arrests, but no convictions or returns to prison noted (N = 146)
d) Parole or mandatory release violation, no return to prison (N = 15)

Unfavorable Category

Any return to prison, or
Any conviction for any new offense, or
Death during the commission of a criminal act, or
Absconding from supervision

[2]See Table 3 for a complete description of predictor variables.

Few would argue against the inclusion of variables pertaining to the present offense (i.e., the offense for which the subject is presently incarcerated). An argument could be made, however, that consideration of criminal history variables may be inappropriate. Such an argument would likely invoke concepts of double-jeopardy. A still more powerful argument potentially could arise over the inclusion of social history variables. Is it just, or legal, to consider variables such as a subject's race or sex, for example, in a paroling decision? Without much effort, the argument readily can be extended to variables such as marital status, educational level, or living arrangements. (Within this framework, the status of the group of variables we have called institutional adjustment could be problematic to the extent that the variables concern institutional adjustment during past, as well as the present, incarceration.)

Tables 11 and 12 summarize regression equations that attend to these variable classifications in their development. Since few would argue that consideration of the present offense would be inappropriate, both present offense variables have been entered (simultaneously) on the first step in calculating both equations. This results in a Multiple R of 0.196.

Subsequent variable groups were included in the following manner. A series of hierarchical step-wise regressions for each variable set was computed. On the first step, all present offense variables were entered. Steps 2 through n (where n exhausts the list of social history variables) allowed social history variables to enter, one at a time, until all variables in the social history group were exhausted. Variables which did not, through this inclusion, add more than 0.003 to the Multiple R^2 were then excluded from the list of the social history variables.

On the next step (Step $n + 1$), all present offense variables and the social history variables selected via the preceding process were allowed to enter simultaneously. The analysis then proceeded, in the same fashion, to exhaust the list of criminal history variables. After the selection of the subset of criminal history variables, we proceeded in like fashion with the institutional adjustment variables. The final result of this process is displayed in Table 11. Table 12 results from the same process, with a different ordering of variable groups (i.e., present offense, followed by criminal history, social history, institutional adjustment).

Both equations result in the same degree of prediction, although they clearly use different variables (and use the different groups of variables differently). Comparison of the two equations demonstrates that: a) including criminal history variables before social history variables increases the Multiple R from 0.20 to 0.34, while b) including social history before criminal history increases R from 0.20 to only 0.30; c) including only five variables, two of which are related to the present offense, and three of which are related to criminal history, allows us to predict almost as well as does any other scheme; and d) the technique outlined above allows, within limits, consideration of both the appropriateness of including variables of different types and their relative impacts on the resulting device.

Validation Study. Multiple-regression techniques, of course, result in a "predicted" outcome score for each individual according to the form of the equation. In operational use, some collapsing scheme (similar in concept to those discussed in an earlier section) must be devised before the equation is used. Accordingly, the "collapsed" regression device is that which should be validated. The scheme used here involves half-standard deviation units on the "predicted" outcome variable.*

As we observed in consideration of the Burgess device, all regression models result in approximately the same degree of prediction, and shrinkage is modest across all devices (Table 13).

Clustering Models

A potential limitation of linear additive models is that they presume homogeneity of the subject population. For example, in calculating regression coefficients we need deal only with the matrix of zero-order correlations among variables across the entire sample. The assumption is made (if implicitly) that the indicated relations hold for population subgroups as well — that is, that the population is in fact homogeneous.

A second limitation may be that linear additive models typically contain no interaction terms. If some particular *combination* of attributes (such as a history of drug addiction and, for example, armed robbery) is a particularly unfavorable prognostic sign, the importance of this "interaction" would be missed in a simple application of regression techniques.

Clustering or configural approaches may in some cases compensate for these limitations in linear additive models, particularly in circumstances where interaction or heterogeneity might be expected.[51] Configural approaches allow for unspecified interactions and heterogeneities which may be present in a population, and they are characterized as *nonlinear* and *hierarchical.*[52]

Predictive Attribute Analysis, and Association Analysis, are the two approaches that have been used most often for criminal justice screening devices, and therefore they have been examined in the present study. Both classify individuals on the basis of the possession or lack of specified attributes, thus permitting development of a typology for risk screening.

Association Analysis

Of central importance to the development of screening devices developed through use of a configural technique are the rules employed for successive partitioning of the sample. In particular, *a priori* guidelines for terminating the analysis must be established. For the Association Analysis reported here, the analysis proceeded by repeatedly subdividing the sample on the attribute "most closely related to all other attributes then present, i.e., that attribute which in a sense is the best single

*We should note that this would typically be expected to reduce explanatory power. For this reason, both "collapsed" and regular coefficients are given in Table 13. Little change results.

Table 11

STEPWISE MULTIPLE REGRESSION: PRESENT OFFENSE, SOCIAL HISTORY, CRIMINAL HISTORY, AND INSTITUTIONAL ADJUSTMENT

Group Variable	Unstandardized Weight	Standardized Weight	R	R²
Present Offense				
Type of Admission	0.0654	0.0575	0.196	0.039
Present Offense	-0.0519	-0.0528	0.196	0.039
Social History				
Opiate Use	0.0491	0.0764	0.300	0.090
Planned Living Arrangement	0.0925	0.0765	0.300	0.090
Employment	0.0360	0.0375	0.300	0.090
Highest Grade Claimed	-0.0355	-0.1213	0.300	0.090
Longest Job in Free Community	-0.0071	-0.0414	0.300	0.090
Criminal History				
Longest Time Free	-0.0355	-0.1213	0.362	0.131
Crime Group	0.0537	0.1320	0.362	0.131
Age at First Arrest	-0.0559	-0.0577	0.362	0.131
Institutional Adjustment				
Prison Punishment	0.1012	0.095	0.373	0.139
Constant	1.4236			
			Validation	
		1972	0.357	0.127

Table 12

STEPWISE MULTIPLE REGRESSION: PRESENT OFFENSE, CRIMINAL HISTORY, SOCIAL HISTORY, AND INSTITUTIONAL ADJUSTMENT

Group Variable	Unstandardized Weight	Standardized Weight	R	R²
Present Offense				
Type of Admission	0.0670	0.0589	0.196	0.039
Present Offense	-0.0460	-0.0468	0.196	0.039
Criminal History				
Longest Time Free	-0.0383	-0.1307	0.335	0.112
Crime Group	0.0600	0.1458	0.335	0.112
Age at First Arrest	-0.0646	-0.0667	0.335	0.112
Social History				
Opiate Use	0.0965	0.0757	0.356	0.127
Living Arrangement	0.0487	0.0846	0.356	0.127
Institutional Adjustment				
Prison Punishment	0.1048	0.0988	0.369	0.136
Constant	1.3957			
			Validation	
		1972	0.358	0.128

'representative' of its fellows for the information they all contain."[53] The statistic used here to assess this relationship was the absolute value of the average intercorrelation of each variable with all other variables. That variable which demonstrates the largest such statistic was, at each step, that on which the sample was partitioned. The process of subdivision was terminated if a) the partition resulted in a subgroup of less than 120 cases, or b) the average absolute intercorrelation was less than 0.10.

Figure 2 presents an example of an analysis that follows the termination rules outlined above. At the first step, the entire sample of 2,382 offenders was partitioned on the "crime group" variable, resulting in one group of 1,021 subjects having at most one prior conviction for forgery, fraud, counterfeiting, burglary, larceny, or auto theft, and one group of 1,361 offenders having two or more such convictions. Each of these samples was further subdivided: the first with respect to age at first arrest, and the second with respect to the nature of the present offense. Further divisions continued in the manner already described until termination, according to the specified rules, was reached.

Although the configuration presented in Figure 2 would not necessarily result in better (or poorer) prediction than would one which terminated earlier in the process, we do lose some degrees of freedom by continuing the process beyond that which may be required. For example, we can see from the ϕ coefficients, which assess relation with the criterion, that some splits may not increase predictive utility. Figure 3 presents an analysis that terminates earlier in the process. Clearly, no reduction in predictive power results.*

As discussed by Wilkins and MacNaughton-Smith, terminal subgroups can be described in terms of attributes associated with the predictor variables which characterize them, and often these characterizations are of considerable importance. The present purpose, however, is simply to examine the predictive utility of the method. As might be expected of a method that relies simply on the creation of increasingly homogeneous subgroups and essentially ignores the criterion measure in its development, the apparent predictive power of the technique is relatively low (Eta = 0.296; MCR = 0.329), although still well within ranges typically found in the development of parole risk-screening devices. As discussed previously, however, one might expect to find less shrinkage on validation precisely for this reason.

Predictive Attribute Analysis

While Association Analysis essentially ignores the outcome criterion in the development of terminal subgroups, Predictive Attribute Analysis successively divides a sample on the single attribute or variable that exhibits the strongest association with the criterion. At each subdivision, the associations within resulting subgroups are measured independently and again divided on that attribute most closely associated with the criterion.

*Given a different result, one might wish to test the significance of the difference between proportions (in favorable outcomes) of adjacent subgroups on the same "limb of the tree," then collapsing (back up the tree) subgroups if and only if the differences are not significant.

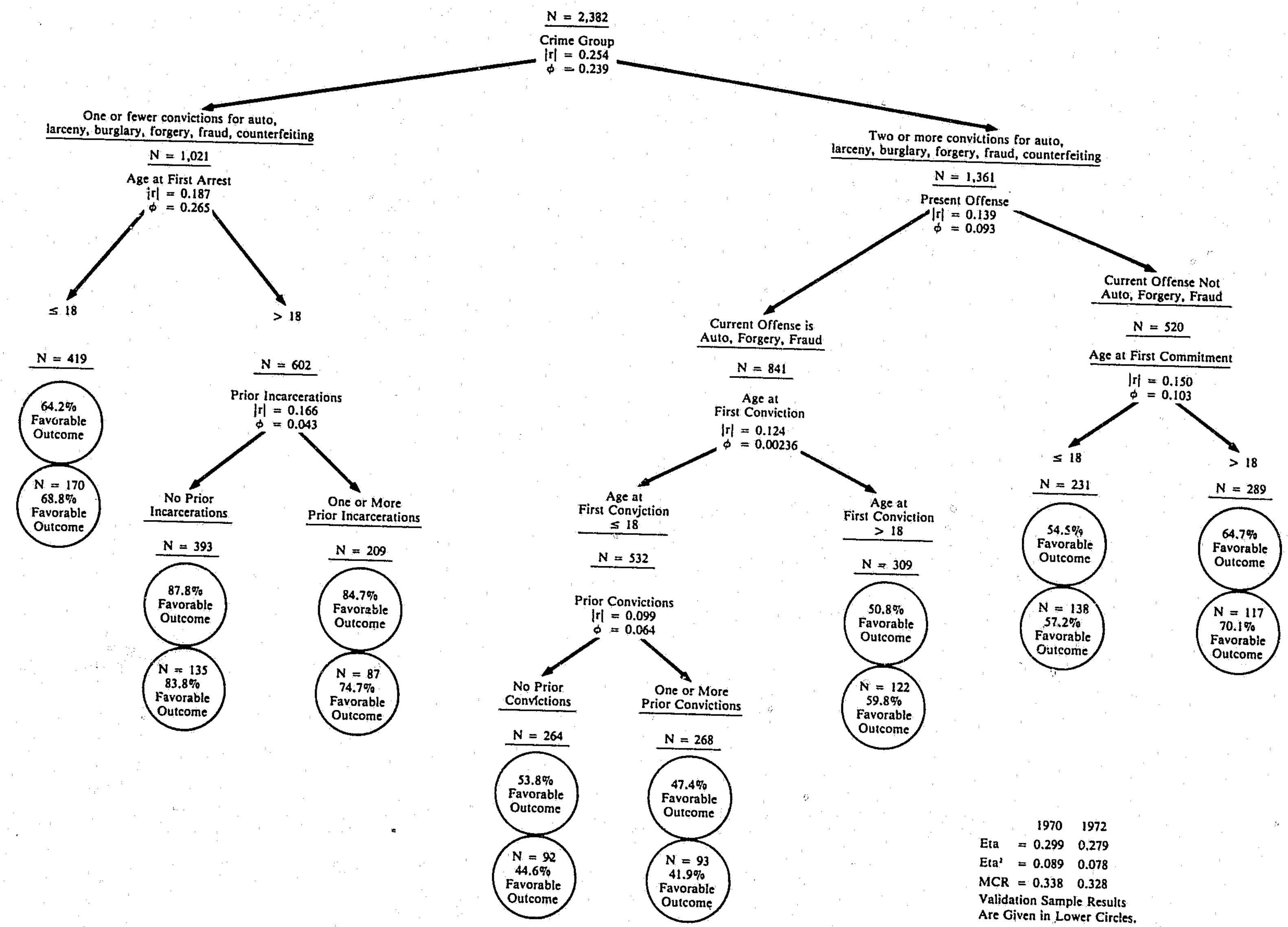

FIGURE 2
ASSOCIATION ANALYSIS

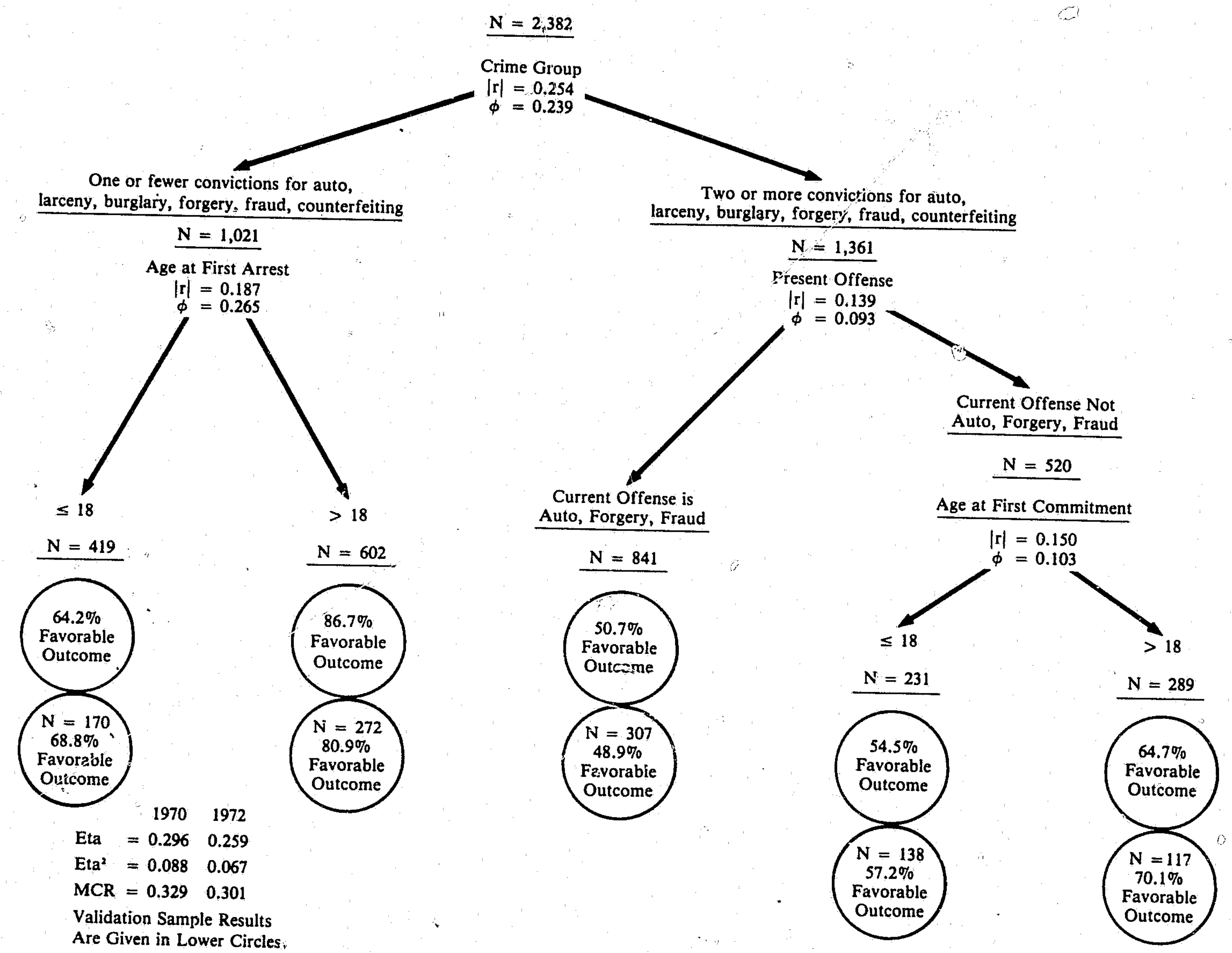

FIGURE 3
ASSOCIATION ANALYSIS WITH FIVE TERMINAL GROUPS

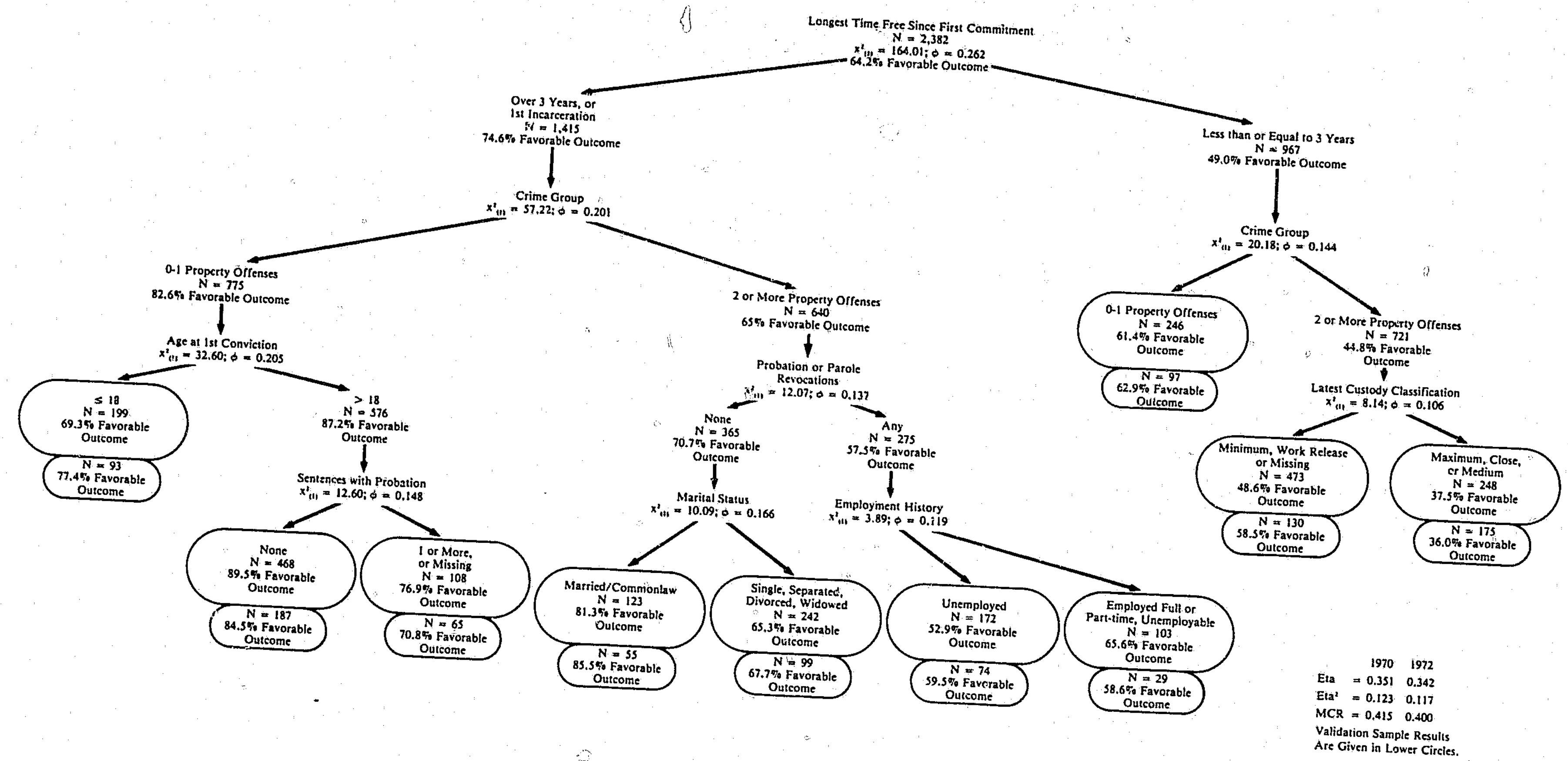

FIGURE 4
PREDICTIVE ATTRIBUTE ANALYSIS WITH TEN TERMINAL GROUPS (29 AVAILABLE PREDICTORS)

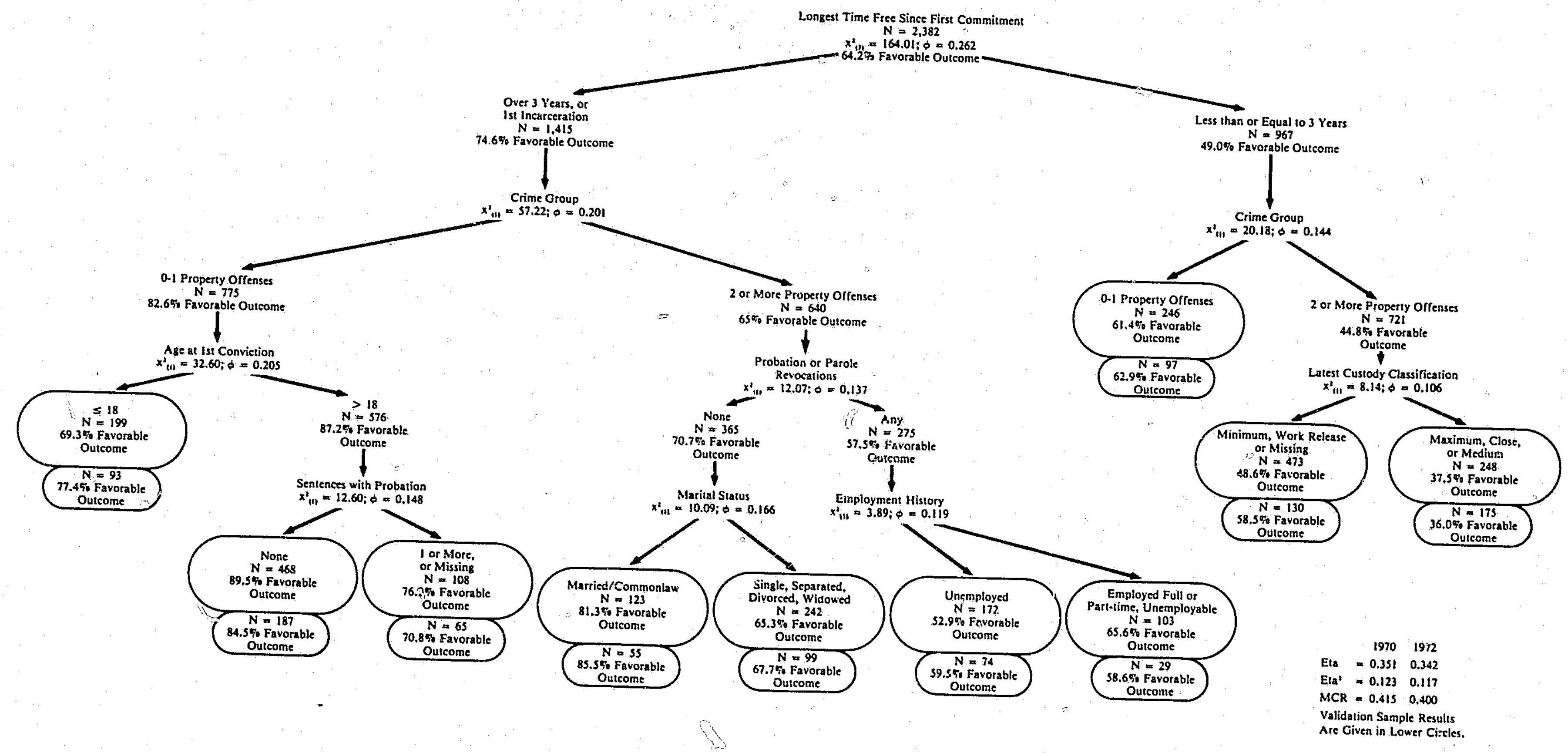

FIGURE 4
PREDICTIVE ATTRIBUTE ANALYSIS WITH TEN TERMINAL GROUPS (29 AVAILABLE PREDICTORS)

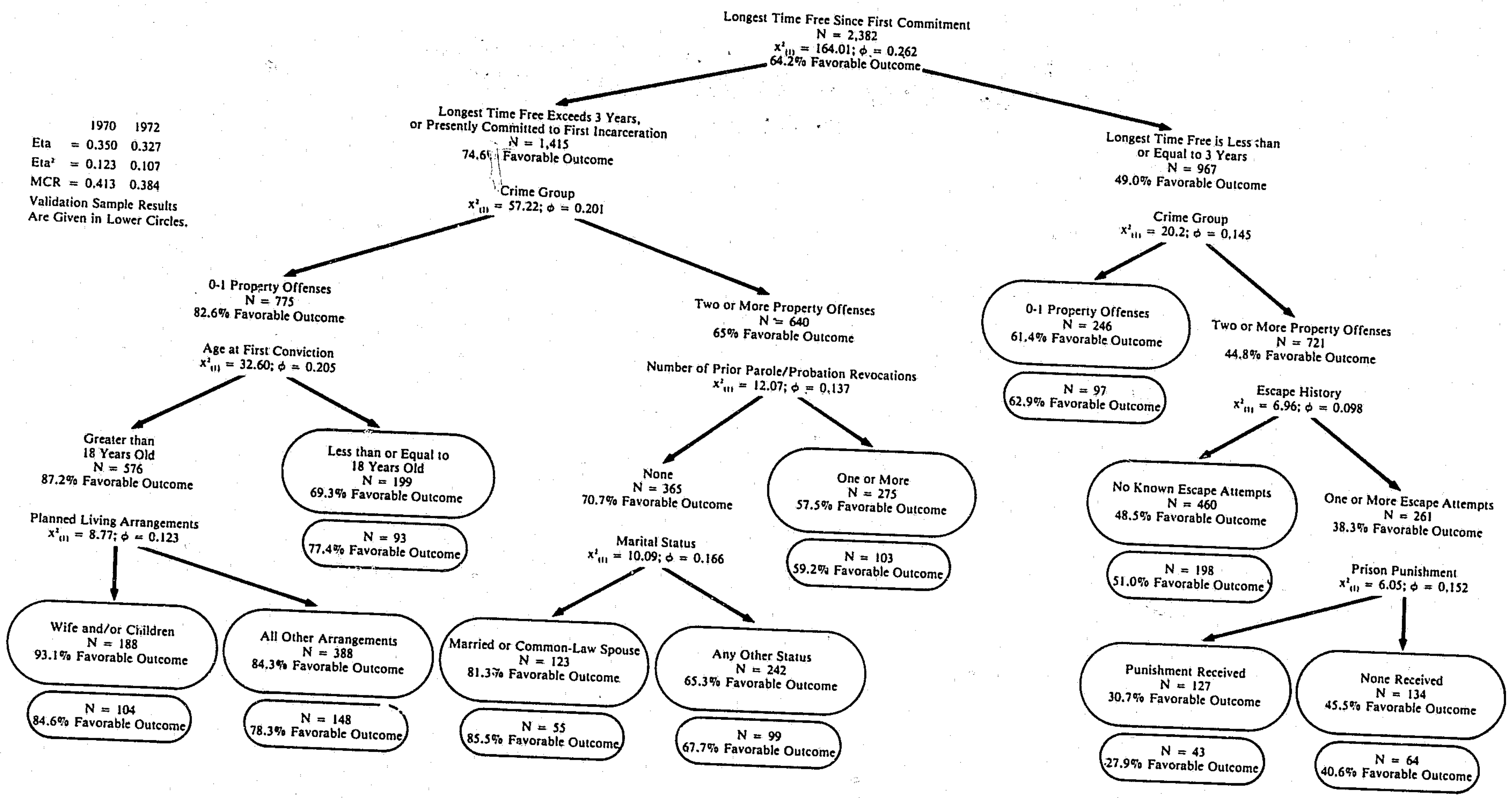

FIGURE 5
PREDICTIVE ATTRIBUTE ANALYSIS WITH TEN TERMINAL GROUPS (19 AVAILABLE PREDICTORS)

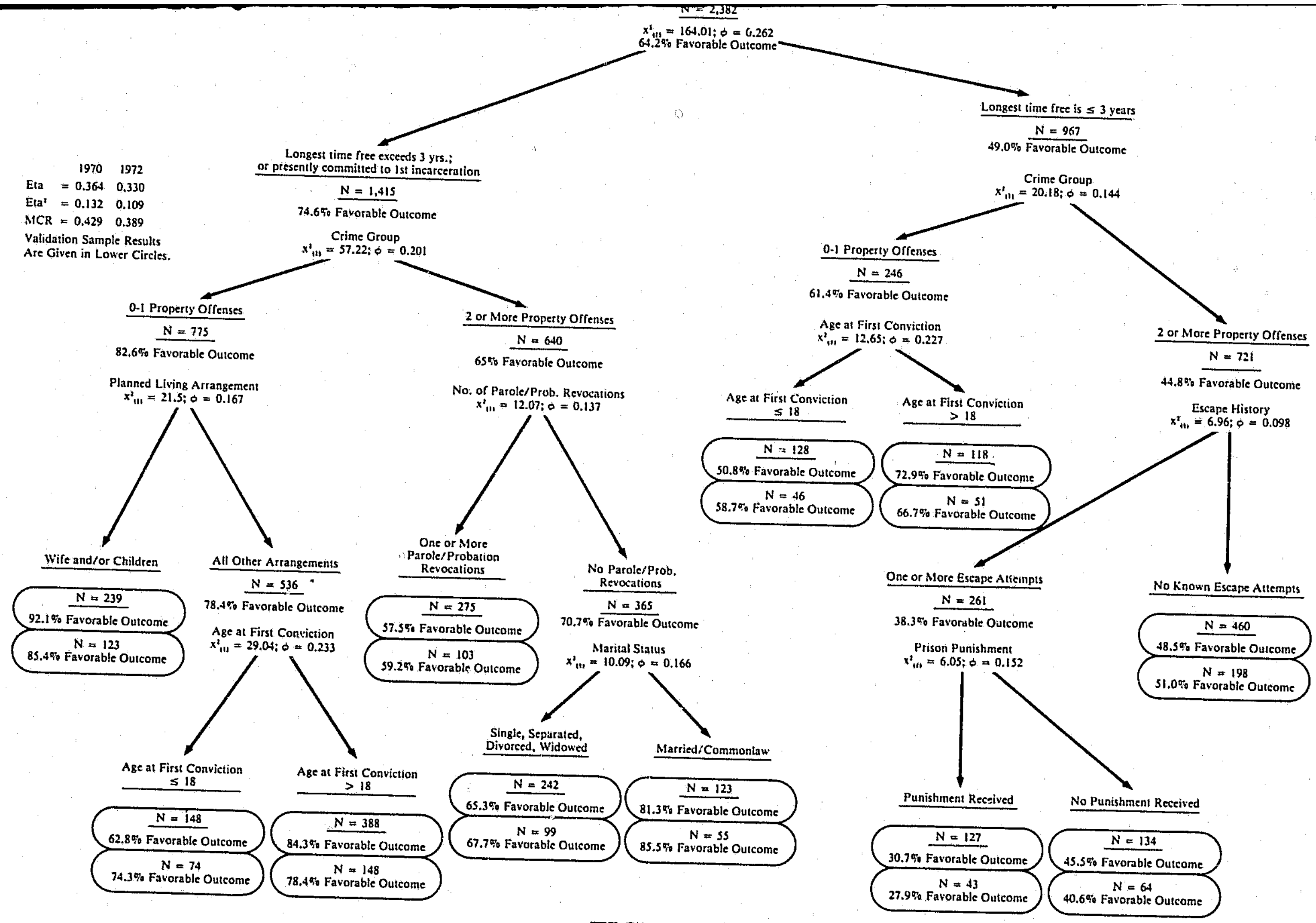

FIGURE 6
PREDICTIVE ATTRIBUTE ANALYSIS
WITH "SECOND BEST OPTION"

Since chi-squared is a function of several parameters and its magnitude is affected (among other things) by the dimensionality of the contingency table and sample size, direct comparisons of the statistic are problematic. Our approach, therefore, was based on the probability level associated with each chi-squared value: that which was lowest was that upon which the sample was subdivided.

The rules for subdivision termination followed for the analyses reported in this section were: a) a split will not be made if it results in a terminal subgroup having fewer than 120 cases, and b) all chi-squared values upon which splits will be made must be statistically significant.

The program used for these analyses also allows the investigator to seek "second-best" alternatives. For example, if the best (i.e., lowest probability) χ^2 results in terminal groups that violate rule a), one can continue through the list until a variable is encountered that meets both criteria, or until the list is exhausted.

Since the variables available for analysis will affect the terminal groups — as well as the paths by which we arrive at them — two variable sets were analyzed: the full set of 29 variables, and those 19 variables found earlier to correlate at least $|0.15|$ with the criterion.

Figure 4 represents an analysis that follows the decision rules specified above but that did not utilize the "second-best" option. All 29 variables were available for inclusion, resulting in ten terminal groups with associated probabilities of success ranging from 0.38 to 0.90. As expected, predictive power was markedly increased over that for the Association Analyses reported above (Eta = 0.351; MCR = 0.415).

Figure 5 gives the results of a similar analysis based on only 19 available variables. Again, ten terminal groups were identified (note, however, that the groups were different). Again, the "second-choice" option was not utilized. The probability of success associated with terminal group membership ranges from 0.31 to 0.93. Despite this apparent modest improvement, no change is reflected in the overall measures of association (Eta = 0.35; MCR = 0.413).

Finally, Figure 6 summarizes an analysis that used 19 available attributes and also allowed the use of the "second-best" option. The solution is slightly more complex (11 terminal groups), but perhaps slightly more powerful (Eta = 0.364; MCR = 0.429).

Association Analysis with Criterion-Referenced Decision Rules

While both configural methods discussed thus far proceed by creating increasingly homogeneous groups, the base for this homogeneity differs markedly between the two. Association Analysis "maximizes" homogeneity across the entire pool of available predictors; Predictive Attribute Analysis, on the other hand, "maximizes" subgroup homogeneity with respect to that single available predictor (on a given choice-node) which is best associated with the criterion. Resulting subgroups are therefore "homogeneous" in both cases, but differently so.

An attractive notion, therefore, is to attempt a simultaneous consideration of both issues. Use of the programs developed for these analyses facilitates this, as the "second-best" option can be included in considerations of Association Analysis as well. All that need be specified are decision rules. Those followed for the analyses presented here were: a) the attribute showing the highest absolute average intercorrelation was that upon which the sample was subdivided, provided that b) the attribute was correlated to at least the extent of 0.10 with the criterion, and c) the split resulted in no subgroup having fewer than 120 cases. The rules were followed in the hierarchical order described, and we have termed the method "quasi-Association Analysis."

Figure 7 gives the resulting structure when 29 available attributes were used, and Figure 8 results from an analysis of 19 available attributes. As expected, the results of both analyses appear to be somewhat more predictive than do those of the standard Association Analyses, but they do not attain the level of the results from Predictive Attribute Analysis.

Validation Studies. The number of validation cases and the associated probability of success for each subgroup for the seven analysis reported above are given (in the lower circles) in Figures 2 through 8, and Table 14 gives a summary of the comparative predictive power of these devices. The shrinkage observed is quite modest across all methods; no single one appeared more (or less) powerful with respect to validation. The Predictive Attribute approach did result in overall better prediction.

Multidimensional Contingency Table Analysis

As already described, the analytic form of the saturated log-linear model given dichotomous variables specifies 2^{n-1} terms, where n is the number of independent variables plus the (in this case, single) dependent variable. Since the model quickly can become unwieldy in the extreme,* the set of potential predictor variables was first reduced to a manageable number.

We decided to restrict attention to dichotomous variables (i.e., attributes) since this dramatically reduces the dimensionality of the matrix, yet, as indicated by our previous analyses, should not dramatically reduce explanatory power. The eight attributes decided upon, given in Table 15, are those which the Predictive Attribute Analysis (Figure 6) showed to successfully cluster observations relative to the outcome criterion. Were we to estimate a saturated model, the right side of the equation would contain 256 (2^8) terms. One of our principal objectives, of course, is to provide a parsimonious model.

The strategy of finding a parsimonious expression of the data in the logit model started with a nine-way cross-tabulation of the outcome variable and eight independent variables, giving 512 cells (2^9) of frequencies. The process of estimating

*That is, use of all 19 potential predictor variables would require forming estimates based on 2^{20} or 1,048,575 cells — roughly 440 times the number of available observations.

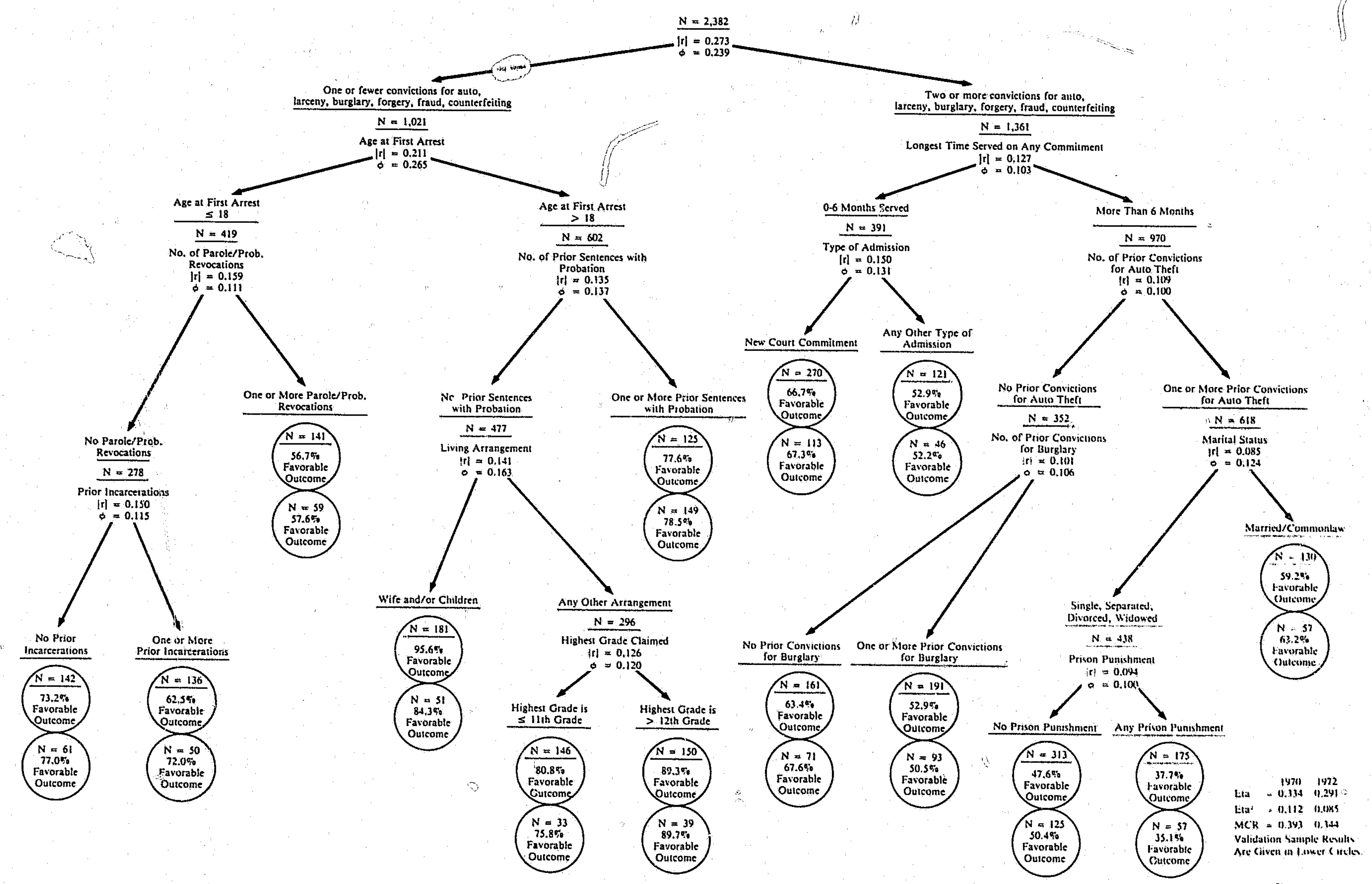

FIGURE 7
QUASI-ASSOCIATION ANALYSIS WITH FOURTEEN TERMINAL GROUPS

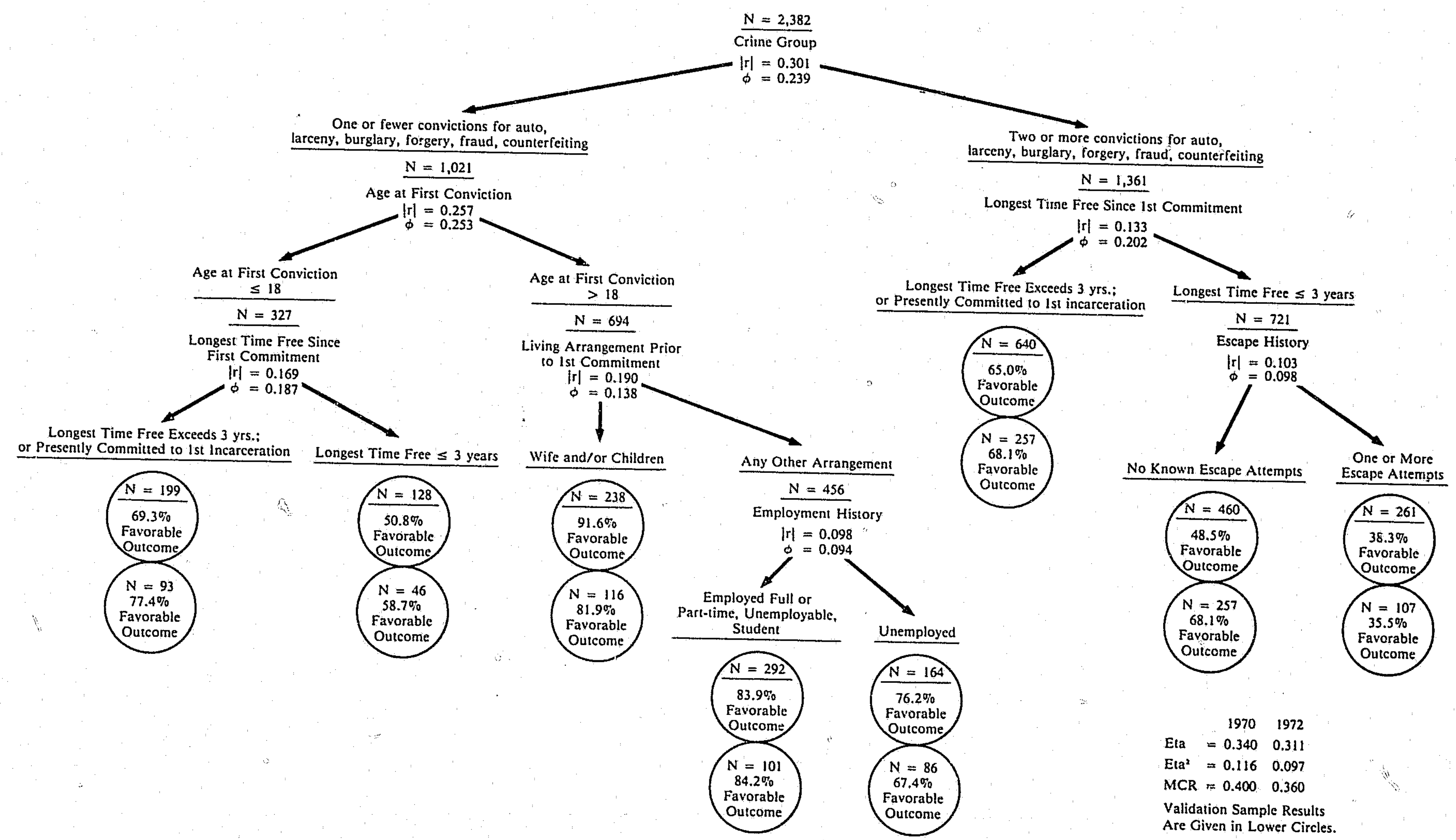

FIGURE 8
QUASI-ASSOCIATION ANALYSIS WITH EIGHT TERMINAL GROUPS

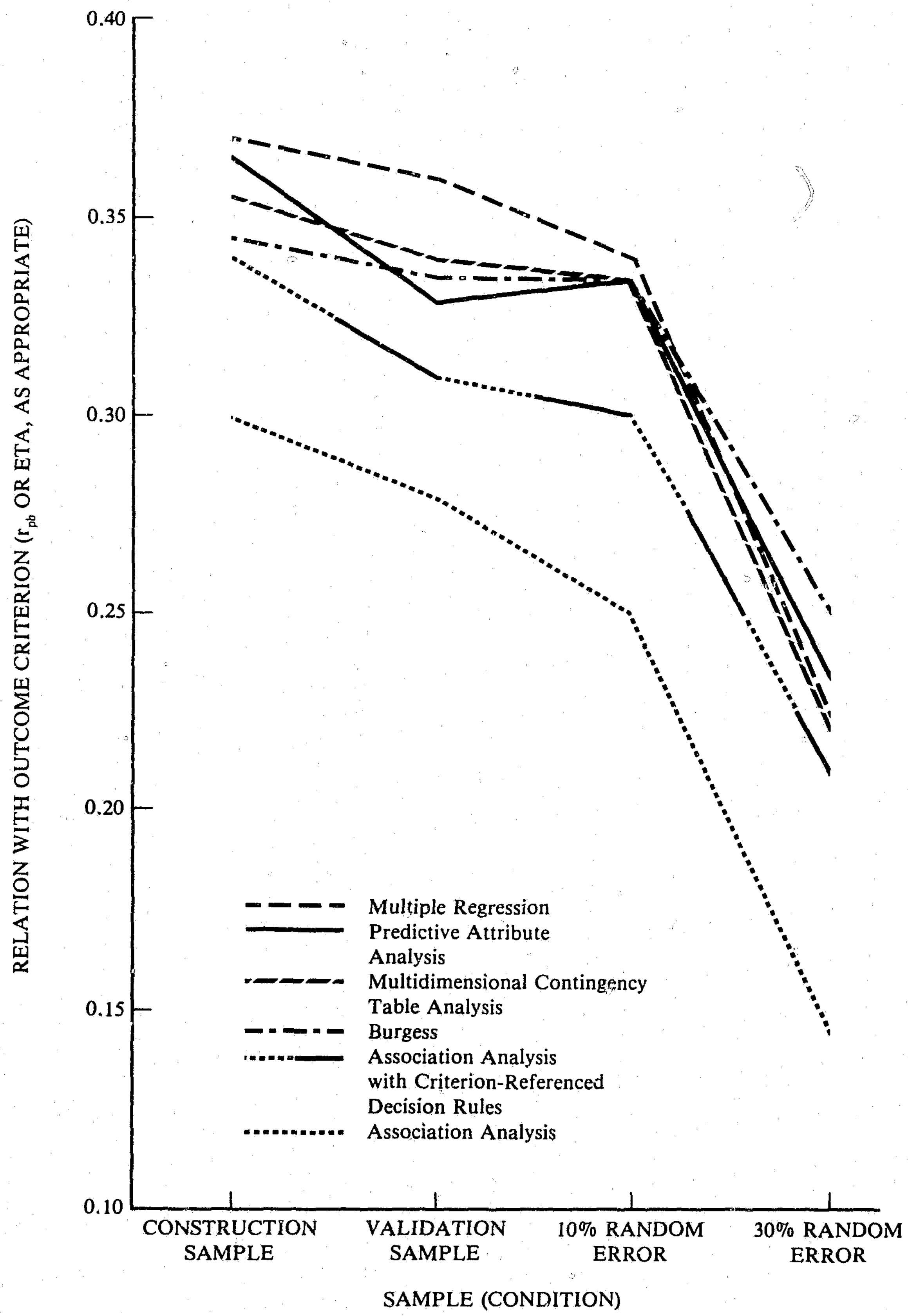

FIGURE 9
ROBUSTNESS OF SIX PREDICTION METHODS

Table 13
MULTIPLE REGRESSION VALIDATION STUDIES

Equation	Multiple Correlation (R) Construction Sample (N = 2,382)	Multiple Correlation (R) Validation Sample (N = 1,004)	Coefficient of Multiple Determination (R^2) Construction Sample (N = 2,382)	Coefficient of Multiple Determination (R^2) Validation Sample (N = 1,004)	Mean Cost Rating (MCR) Construction Sample (N = 2,382)	Mean Cost Rating (MCR) Validation Sample (N = 1,004)	Shrinkage R^2	Shrinkage MCR
***Eighteen Predictor* Variables**	0.382 (0.385)	30.356 (0.357)	0.146 (0.148)	0.127 (0.128)	0.458	0.423	0.019 (0.020)	30.035
Stepwise Procedure: Present Offense, Social History, Criminal History, Institutional Adjustment	0.370 (0.373)	0.362 (0.357)	0.137 (0.139)	0.131 (0.127)	0.440	0.436	0.006 (0.012)	0.024
Stepwise Procedure: Present Offense, Criminal History, Social History, Institutional Adjustment	0.366 (0.369)	0.355 (0.358)	0.134 (0.136)	0.126 (0.128)	0.436	0.439	0.008 (0.012)	—

Note: Values in parentheses reflect relation with the outcome variable in uncollapsed form. For definition of the dependent variable (outcome), see footnote 1, Table 10.

Table 14
CONFIGURAL APPROACH VALIDATION STUDIES

Configural Solution	Coefficient of Non-Linear Correlation (Eta) Construction Sample	Coefficient of Non-Linear Correlation (Eta) Validation Sample	Eta^2 Construction Sample	Eta^2 Validation Sample	Mean Cost Rating (MCR) Construction Sample	Mean Cost Rating (MCR) Validation Sample	Shrinkage Eta^2	Shrinkage MCR
Association Analyses:								
Eight Terminal Groups	0.299	0.279	0.089	0.078	0.338	0.328	0.011	0.010
Five Terminal Groups	0.296	0.259	0.088	0.067	0.329	0.301	0.021	0.028
Predictive Attribute Analyses:								
Ten Terminal Groups (30 Available Predictors)	0.351	0.342	0.[illegible]	0.117	0.415	0.400	0.006	0.015
Ten Terminal Groups (19 Available Predictors)	0.350	0.327	0.123	0.107	0.413	0.384	0.016	0.029
Eleven Terminal Groups (19 Available Predictors, "Second-Best" Option Utilized)	0.364	0.330	0.132	0.109	0.429	0.389	0.023	0.040
Association Analysis with Criterion-Referenced Decision Rules:								
Fourteen Terminal Groups (30 Available Predictors)	0.334	0.291	0.112	0.085	0.393	0.344	0.027	0.049
Eight Terminal Groups (19 Available Predictors)	0.340	0.311	0.116	0.097	0.400	0.360	0.019	0.040

Table 15
MULTIDIMENSIONAL CONTINGENCY ANALYSES: INDEPENDENT VARIABLES

Variable Description	Name	Letter Code
Longest Time Free Between Incarcerations: 3 or more years (or first offender) vs. less than 3 years	TIMEFREE	("T")
Property Crime History: No or one property offense vs. two or more	PROPCRIM	("C")
Planned Living Arrangement: Plans to live with spouse and/or children vs. all other planned living arrangements	PLIVARR	("L")
Age at First Conviction: 19 or older vs. less than 19	AGECON	("A")
Number of Paroles and/or Probations Revoked: Never vs. once or more	PARREV	("R")
Marital Status: Married or common-law arrangement vs. all others	MARITAL	("M")
Escape History: No attempts vs. one or more	ESCAPE	("E")
Punishment in Prison: None vs. once or more	PRISPUN	("P")

an "optimal" model involved the determination of which, if any, of these eight predictors could be dropped from the model (i.e., which of the $\gamma^{A}{}_{i} = 1$). After dropping any of the attributes, a cross-tabulation of those remaining was used to see which interactions, if any, should be added. Using the letters noted in Table 15 to identify each of the predictors, a "main effects only" model was estimated:

$$\Omega = \gamma\, \gamma^{T}{}_{i}\, \gamma^{C}{}_{j}\, \gamma^{L}{}_{k}\, \gamma^{A}{}_{l}\, \gamma^{R}{}_{m}\, \gamma^{M}{}_{n}\, \gamma^{E}{}_{o}\, \gamma^{P}{}_{p}$$

The significant results of this estimation are displayed in Table 16, line 2. Note that the log-likelihood χ^2 estimate for this model is 182.29.

Reference to a table of chi-squared values yields a probability value greater than 0.5, indicating a good fit between the observed and expected values. Use of such probabilities exclusively, however, is sub-optimal, since the algorithm used to estimate the model creates maximum likelihood estimates of effects on its right side, which in turn yield a *minimized* value of the log likelihood chi-squared. This pro-

vides a statistic that will yield a conservative test for rejection of the "null" hypothesis (hypothesis H_2 in this case), but not necessarily a conservative test for accepting that hypothesis.*

Table 16
MAIN EFFECTS MODELS FROM THE 2⁸ CONTINGENCY TABLE

H_i	Descriptions	χ^2 (d.f.)	$\chi^2_1 - \chi^2_i$ (p)	R^2_i	$\chi^2_i - \chi^2_2$ (p)	$r^2_{2.i}$
H_1	No Independent Effects	489.48 (255)	— —	—	— —	—
H_2	All 8 Main Effects	189.29 (247)	307.19 (<0.005)	0.63	— —	—
H_3	H_2, except TIMEFREE	223.08 (248)	266.40 (<0.005)	0.54	40.79 (<0.005)	0.18
H_4	H_2, except PROPCRIM	213.81 (248)	275.67 (<0.005)	0.56	31.52 (<0.005)	0.15
H_5	H_2, except PLIVARR	185.73 (248)	303.75 (<0.005)	0.62	3.44 (<0.05)	0.02
H_6	H_2, except AGECON	187.33 (248)	302.15 (<0.005)	0.62	5.04 (<0.01)	0.03
H_7	H_2, except PARREV	200.90 (248)	288.58 (<0.005)	0.59	18.61 (<0.005)	0.09
H_8	H_2, except MARITAL	188.24 (248)	301.24 (<0.005)	0.62	5.95 (<0.01)	0.03
H_9	H_2, except ESCAPE	192.81 (248)	296.67 (<0.005)	0.61	10.52 (<0.005)	0.05
H_{10}	H_2, except PRISPUN	202.75 (248)	286.73 (<0.005)	0.59	20.46 (<0.005)	0.10

Since we are more interested in the possibility of accepting this model, another view of "goodness of fit" is wanted, which would yield a better grasp of how much unexplained error a chi-squared of 182.29 really represents. We therefore followed the suggestion of Goodman[54] by first finding the difference between the error in this model and the error that would obtain if no main effects were postulated (model H_1 in the table). This number (307.19 in the table) is distributed as chi-squared with (v_1 - v_2) degrees of freedom (where H_1 and H_2 have v_1 and v_2 degrees of freedom, respectively). In our case, a probability of less than 0.001 indicates an extremely low risk of error in rejecting the hypothesis that there are no main effects. Goodman, however, suggests a statistic, which he calls R^2, that is obtained by dividing the decrease in chi-squared in Model 2 vs. Model 1, by the value of chi-squared in Model 1. This ratio,

*It should be noted that the null hypothesis in this case is the hypothesis *of the model:* hence, failure to reject indicates a good fit between the model (expected) and the data (observed). In the classical case, of course, the null hypothesis typically states independence.

Table 17
LOGIT MODELS FROM THE 2^7 CONTINGENCY TABLE

H_i	Descriptions	χ^2 (d.f.)	$\chi_1^2 - \chi_i^2$ (p)	R_i^2	$\chi_2^2 - \chi_i^2$ (p)	$r^2_{2.i}$
H_1	Mean Effect Only	375.65 (63)	312.65 (<0.005)	—	—	—
H_2	Main Effect Only	65.60 (57)	310.05 (<0.005)	0.83	—	—
H_3	H_2 + (TIMEFREE x PROPCRIM)	64.90 (56)	310.75 (<0.005)	0.83	0.70 (<0.25)	0.01
H_4	H_2 + (TIMEFREE x PARREV)	61.88 (56)	313.77 (<0.005)	0.84	3.72 (<0.05)	0.06
H_5	H_2 + (TIMEFREE x MARITAL	65.45 (56)	310.20 (<0.005)	0.83	0.15 (<0.25)	0.00
H_6	H_2 + (TIMEFREE x ESCAPE)	65.23 (56)	310.42 (<0.005)	0.83	0.37 (<0.25)	0.01
H_7	H_2 + (TIMEFREE x PRISPUN)	65.34 (56)	310.31 (<0.005)	0.83	0.26 (<0.25)	0.00
H_8	H_2 + (PROPCRIM x PARREV)	64.10 (56)	311.55 (<0.005)	0.83	1.50 (<0.25)	0.02
H_9	H_2 + (PROPCRIM x MARITAL)	60.87 (56)	314.78 (<0.005)	0.84	4.73 (<0.025)	0.07
H_{10}	H_2 + (PROPCRIM x ESCAPE)	64.47 (56)	311.18 (<0.005)	0.83	1.13 (<0.25)	0.02
H_{11}	H_2 + (PROPCRIM x PRISPUN)	59.96 (56)	315.69 (<0.005)	0.84	5.64 (<0.01)	0.09
H_{12}	H_2 + (PARREV x MARITAL)	64.26 (56)	311.39 (<0.005)	0.83	1.34 (<0.10)	0.02
H_{13}	H_2 + (PARREV x ESCAPE)	63.70 (56)	311.95 (<0.005)	0.83	1.90 (<0.10)	0.03
H_{14}	H_2 + (PARREV x PRISPUN)	63.03 (56)	312.62 (<0.005)	0.83	2.57 (<0.01)	0.04
H_{15}	H_2 + (MARITAL x ESCAPE)	63.16 (56)	312.49 (<0.005)	0.83	2.44 (<0.10)	0.04
H_{16}	H_2 + (MARITAL x PRISPUN)	63.42 (56)	312.23 (<0.005)	0.83	2.18 (<0.10)	0.03
H_{17}	H_2 + (ESCAPE x PRISPUN)	65.55 (56)	310.10 (<0.005)	0.83	0.05 (<0.25)	0.00
H_{18}	H_2 + (TIMEFREE x PARREV) + (PROPCRIM x MARITAL)	61.02 (55)	314.63 (<0.005)	0.84	4.58 (<0.025)	0.07
H_{19}	H_2 + (TIMEFREE x PARREV) + (PROPCRIM x PRISPUN)	56.54 (55)	319.11 (<0.005)	0.85	9.06 (<0.01)	0.14
H_{20}	H_2 + (PROPCRIM x MARITAL) + (PROPCRIM x PRISPUN)	59.41 (55)	316.24 (<0.005)	0.84	6.19 (<0.025)	0.09
H_{21}	H_2 + (TIMEFREE x PARREV) + (PROPCRIM x MARITAL) + (PROPCRIM x PRISPUN)	56.15 (54)	319.50 (<0.005)	0.85	9.45 (<0.005)	0.14

Table 18
PARAMETERS OF "BEST" LOGIT MODELS

Variable	$\gamma(H_2)$	$\gamma(H_{11})$
Constant	1.65	1.62
TIMEFREE	1.38	1.38
PROPCRIM	1.38	1.32
PARREV	1.23	1.23
MARITAL	1.23	1.25
ESCAPE	1.18	1.18
PRISPUN	1.26	1.30
Interaction Effects (PROPCRIM × PRISPUN)	—	1.13

Table 19
A COMPARISON OF THE PREDICTIVE UTILITY OF SIX METHODS

Device (Method Used)	Measure of Association (r_{pb}; R; Eta)		Proportion of Outcome Variance Explained		Mean Cost Rating (MCR)		Shrinkage	
	Construction Sample (N = 2,382)	Validation Sample (N = 1,004)	Construction Sample (N = 2,382)	Validation Sample (N = 1,004)	Construction Sample (N = 2,382)	Validation Sample (N = 1,004)	Proportion of Outcome Variance Explained	MCR
Burgess: 19 Items:								
(a) Uncollapsed version	0.350	0.333	0.123	0.111	0.429	0.423	0.012	0.006
(b) Operationally collapsed (½ standard deviation units)	0.345	0.335	0.119	0.112	0.408	0.404	0.007	0.004
Multiple Regression: Hierarchical Inclusion; Present Offense Variables, Criminal History Variables, Institutional Adjustment Variables:								
(a) Uncollapsed version	0.373	0.357	0.139	0.127	0.440[a]	0.436[a]	0.012	0.024
(b) Operationally collapsed (half-standard deviation units)	0.370	0.362	0.137	0.131	0.440	0.436	0.006	0.024
Association Analysis: Eight Terminal Groups	0.299	0.279	0.089	0.078	0.338	0.328	0.011	0.010
Predictive Attribute Analysis: Eleven Terminal Groups, "Second Best Option" Used	0.364	0.330	0.132	0.109	0.429	0.389	0.023	0.040
Association Analysis with Criterion-Referenced Decision Rules: Eight Terminal Groups	0.340	0.311	0.116	0.097	0.400	0.360	0.019	0.040
Multidimensional Contingency Table Analysis:								
(a) 64-cell solution	0.388	0.391	0.151	0.153	0.462	0.457	—	0.005
(b) Operationally collapsed version (seven probability levels)	0.355	0.339	0.126	0.115	0.419	0.394	0.011	0.025

[a]Based on operationally collapsed instrument.

$\left(\frac{\chi_1^2 - \chi_2^2}{\chi_1^2}\right)$, is a measure of proportional reduction in error. That is, Model H_2 reduces the error (as measured by chi-squared) in H_1 by 63 percent.*

Since we seek parsimony as well as error reduction, we next analyzed the importance of each of the main effects in turn, looking for candidates to drop. Thus, eight more models were postulated, each with all but one of the eight independent effects of H_2. The results are displayed as hypotheses H_3 to H_{10}. For each hypothesis H_i, subtracting the chi-squared in H_2 from χ^2 of H_i gives the reduction in error attributable to introduction of the effect missing in H_i and present in H_2. Again, each of these statistics is distributed as chi-squared with $v_i - v_2$ degrees of freedom. Examination of these probabilities alone would lead to rejection of only one of these effects, PLIVARR, at the 5 percent confidence level. We again wished to look, however, not only at the significance probabilities of these effects, but also at the proportional reduction in error attributable to them. We again followed Goodman, who defines the coefficient of partial correlation as $r^2 w H_2 \cdot H_i = \frac{\chi_1^2 - \chi_2^2}{\chi_1^2}$, i.e., the proportional reduction in error, given H_i, that results from H_2. Examination of the partial r^2 column in Table 16 shows three general levels of the statistic: TIMEFREE and PROPCRIM at the highest level (0.18 and 0.15 respectively); PARREV and PRISPUN in the middle (0.09 and 0.10); and PLIVARR, AGECON, MARITAL, and ESCAPE at the lowest level (0.02, 0.03, 0.04, and 0.05). Each of these last four might be considered as candidates to be dropped. The questions are, of course, how many, and which?

In order to resolve this problem, 11 more models were computed, in which all possible pairs and triples and the only quadruple of these four effects were dropped simultaneously. Results are given in Table A-22. All of the sets of effects are shown to be statistically significant. Model H_{11} (PLIVARR and AGECON dropped) retains, however, the highest R^2 and, correspondingly, the lowest partial r^2. Consequently, it was decided that all further model estimation would be restricted to effects involving TIMEFREE, PROPCRIM, PARREV, MARITAL, ESCAPE, and PRISPUN only.

*Although Goodman uses R^2 to symbolize proportional reduction of error, it should not be confused with the coefficient of determination derived from multiple regression, even though that also is conventionally labeled R^2. There are two main differences in the definition of these statistics: 1) the basic observation units over which total sample error is aggregated, and 2) the measure of error used. For Goodman's R^2, the basic unit of observation is a single cell of a contingency table. In the case of regression, at least in this study, individual cases are the units over which error is aggregated. The difference in error measurement is as follows: regression predicts a "score" (in our case, an estimated probability of success) from which is subtracted the observed score (0 for failure, 1 for success). This difference is squared and aggregated over all cases. In the case of logit analysis, for each combination of independent variables there are observed frequencies of successes and failures, and expected frequencies. The (natural) logarithm of the ratio of expected counts to observed counts is multiplied by the observed count, aggregated over all cells and multiplied by 2. In part because it is easier to predict the overall probability of success for a group of cases than it is to predict success or failure on a case-by-case basis, the R^2s defined by Goodman appear larger than those derived under regression analysis. For example, the probability of success under the logit model was calculated for each individual case and correlated with observed outcomes, yielding correlation coefficients of 0.337 (1970) and 0.336 (1972), numbers remarkably similar to the regression results and considerably less than the similar statistics calculated according to the Goodman suggestion.

Table 17 displays all the models. Note that Model H_1 (no main effects) and Model H_2 (all *six* main effects) were calculated again, yielding different chi-squared estimates (and R^2s) from those in Table 16. This is due to the reduction of the size of the contingency table to 128 cells (as a result of collapsing the earlier table over the variables PLIVARR and AGECON). R^2s in the 0.83 to 0.85 range are now obtained. Given that we have settled on six main effects, the next step is to identify whether there are any interactions that should be added to the model. Thus, 15 models (hypotheses H_3 through H_{17} in Table 17) were calculated, each model representing model H_2 plus one two-way interaction effect. None of these raised the overall R^2 by more than 0.01, although three of them (H_4, H_9, and H_{11}) demonstrated significant interaction effects at the 5 percent level of confidence. The significance and partial r^2s of these three models, however, were not viewed as important enough to include in the model, given the consequent reduction in parsimony. For completeness, four additional models were computed, including all combinations of these two-way interaction effects (Models H_{18} - H_{21}). Again the results were statistically significant, but the methods did not add greatly to R^2. It is at this point that evaluating the trade-off between parsimony and reducing prediction error seems to be most clear. A preference for parsimony may lead one to Model H_2; whereas, a preference for reduction of error would point to Model H_{11}. The parameters for these two models are displayed in Table 18.

Validation Study. As already noted, the models in Table 18 have six dichotomous independent variables. These yield 64 (2^6) odds ratios (and 64 success probabilities), obviously much too large a set to be used for practical prediction purposes. (Recall the similar case of the Burgess and regression instruments.) Examination of the predicted probabilities and the number of cases in each cell, however, yielded a convenient aggregation of cells into seven ranges of predicted success probabilities (with cut points of 0.385, 0.455, 0.858, 0.665, 0.775, and 0.845). Categorizing all of the 1970 observations into the appropriate one of these seven "cell groups" yields an Eta of 0.355, an MCR statistic of 0.419, and a Pearson correlation coefficient of 0.349. (The original 64 cell categories, of which two were empty, had Eta and MCR of 0.388 and 0.462, respectively. Since the original cells were not ordered, a Pearson correlation coefficient was not calculated.)

Although Goodman's R^2 was not used directly to compare power of the logit model with that of the other instruments, it can be used as a validation statistic. Application of Model H_2 (our preference was for parsimony) to the 1972 data yielded a chi-squared of 36.31. When compared to the chi-squared which would occur due to the (1972) "base rate" only, an R^2 of 0.76 was obtained (compared to 0.83 for the construction sample), resulting in moderate shrinkage (0.07).

5 Summary and Discussion

The development and validation of statistical risk-screening devices, while relatively straightforward, involve a number of practical and methodological issues. Even a method as apparently simple as that of Burgess requires careful attention to several problems: how many items should be included? in what form? how should the bivariate (score/criterion) distribution be collapsed to provide an operationally useful prediction instrument? As stressed in an earlier section, many of these issues involve policy decisions as well as simple statistical decisions. No technique is likely to result in an operationally useful, reliable, and valid decisionmaking aid without careful attention to these kinds of issues.

Results presented thus far have demonstrated the construction of several kinds of devices. We turn now to a consideration of their relative power.

Several of the methods (Burgess, multiple regression, and the log-linear model specifically) do not provide a ready-made "decisionmaking instrument"; rather, one must be constructed from the results of developing and testing the model. The Burgess technique provides merely a description of a bivariate distribution—which may not provide even a monotonic increasing function (in relation to proportions in the favorable outcome classification). The regression method, which must result in an equation specifying a monotonic ordering, also defines a continuous (outcome) variate. Finally, the log-linear model results (typically) in a large set of non-ordered contingency cells. An operational "instrument" must be developed from each.

Accordingly, the comparison of methods given in Table 19 validates instruments rather than equations or models (although the latter coefficients are given for completeness).

The evidence presented in Table 19 suggests no clear advantage of any given method. Prediction using all methods of instrument development tested is at best modest, although: a) prediction is better than that which would result from the simple use of the base-rate alone (regardless of the method of construction employed), and b) the estimates derived here are well within ranges typically found in "state-of-the-art" studies of recidivism prediction.

With the possible exception of the Association Analyses (the predictive validities of which are depressed relative to the remainder), all techniques result in virtually the same degree of predictive efficiency. On the basis of these analyses, at least, those who would develop risk-screening devices for operational use would be advised to base their decisions as to the method(s) to employ on factors other than the statistical power inherent in the methods considered. Indeed, the practical and statistical simplicity of the Burgess method commends its use in many applications.

Limitations to Analyses

Having made this bold proclamation, we hasten to add a few caveats. In some respects, the data used — while the best known to be available — are not optimal for a study that purports to compare the relative power of different predictive methods.

Table 20
A COMPARISON OF SIX PREDICTION METHODS UNDER VARYING DEGREES OF KNOWN RELIABILITY

	Original Construction Sample (N = 2,382)		Validation Sample (N = 1,004)		Construction Sample with 10% Random Error (N = 2,382)		Construction Sample with 30% Random Error (N = 2,382)	
Device (Method Used)	Association with Outcome (r or Eta)	MCR	Association with Outcome (r or Eta)	MCR	Association with Outcome (r or Eta)	MCR	Association with Outcome (r or Eta)	MCR
Burgess: 19 items, half-standard deviation collapsing scheme	0.345	0.408	0.335	0.404	0.334	0.395	0.249	0.295
Multiple Regression: Stepwise procedure, half-standard deviation collapsing scheme	0.370	0.440	0.362	0.436	0.339	0.402	0.222	0.264
Association Analysis: Eight terminal groups	0.299	0.338	0.279	0.328	0.247	0.285	0.145	0.167
Predictive Attribute Analysis: 19 available predictors, eleven terminal groups	0.364	0.429	0.330	0.389	0.338	0.399	0.233	0.273
Association Analysis with Criterion-Referenced Decision Rules: 19 available predictors, eight terminal groups	0.340	0.400	0.311	0.360	0.305	0.358	0.210	0.245
Multidimensional Contingency Table Analysis: Seven probability groupings	0.355	0.419	0.339	0.394	0.338	0.399	0.224	0.252

Table 21
INTERCORRELATIONS OF SIX PREDICTION INSTRUMENTS

	Burgess	Multiple Regression	Association Analysis	Predictive Attributes Analysis	Quasi-Association Analysis	Multidimensional Contingency Table Analysis
Burgess Device (19 items; half-standard deviation unit Operational Collapse)	—	0.864 (0.877)	0.782 (0.777)	0.833 (0.830)	0.816 (0.803)	0.824 (0.830)
Multiple Regression (Stepwise inclusion; half-standard deviation unit Operational Collapse)		—	0.717 (0.717)	0.818 (0.813)	0.784 (0.769)	0.848 (0.838)
Association Analysis (Eight Terminal Groups)			—	0.736 (0.745)	0.781 (0.782)	0.640 (0.645)
Predictive Attribute Analysis (Eleven Terminal Groups; second-best option utilized)				—	0.912 (0.905)	0.866 (0.867)
Quasi-Association Analysis (Eight Terminal Groups)					—	0.834 (0.645)
Multidimensional Contingency Table Analysis (Operationally Collapsed)						—

Note: Values in parentheses are based on the 1972 validation sample.

For example, with respect to the regression model, the data base includes few items having the requisite level of measurement (i.e., interval or ratio), thus partially robbing the technique of potential power. Using the log-linear model, we had to limit the number of categories for any given variable to two, given the sizes of the samples

available. This too could restrict the potential power (and hence potential benefits) of the method.*

The clustering methods employed are both divisive; one might well suspect that an agglomerative algorithm (which typically would make use of a substantially larger number of data items) could perform better.

As stressed early in this report, the criterion chosen for examination is critical. Studies of this type typically employ a criterion such as that decided upon here—that is, a simple dichotomous "good/bad," "success/failure" classification. Such a decision has serious statistical implications. Restriction of range constrains coefficients of relation.[55] Given a more sophisticated outcome measure, our results could well have been different.

The points made earlier in our discussion of policy issues still apply. If the purpose is the development of an operationally viable aid to practical decisionmaking, then the criteria that decisionmakers are willing to accept (and employ) may be those that should be employed.† This is not to suggest that some alternative criterion measures may not be viable and employable, but simply that they have not typically been employed nor has their viability been assessed. In fact, had a different (non-dichotomous) criterion been used, other techniques (such as multiple discriminant function analysis) could have been examined. If, for example, one merely classified outcomes into three sets—1) return to prison with a new felony conviction, 2) any other return to prison, and 3) no return to prison — then the discriminant function (which is readily generalized to more than two groups) could have been used. (N-1 equations would then be derived, where N is the number of criterion classifications.) Further, one may continue to hope that progress may be made toward a more adequate scaling of what is meant by relatively favorable/unfavorable behavior after release from prison.

*It is not entirely clear, of course, whether this is a "fault" of the method or of the data — the point being that a high-powered sports car and an economy import both may be constrained by a 55 mph speed limit.

†We have argued elsewhere (Gottfredson, D., *op cit.*, 1967) that improved measures of recidivism are seriously needed.

6 Which Actuarial Approach — Revisited

Results of these studies suggest no clear-cut advantage to any of the methods examined. The results summarized in Table 19 suggest that all methods used provide essentially equivalent predictive utility, with the possible exception of the Association Analysis—the primary purpose of which is not, of course, predictive. It is important to stress, however, that all analyses performed were constrained to a greater or lesser degree by the available data. The reader should therefore bear in mind that we make no claim other than that, given the nature of the data and of the criteria typically available for similar prediction studies, the choice of prediction methods (i.e., statistical techniques for the development of operationally useful decisionmaking aids) would seem to make little difference.

The data used in this series of studies—while by no means optimal—are believed to be similar to the best currently available. A large pool of potential predictor variables was available, and the reliability (of coding information from base files—an exceedingly difficult and time-consuming task) of this information is known to be quite acceptable.

Typically, however, those who would develop statistical aids to decisionmaking are faced with a different situation. In particular, they are likely to be confronted with data of unknown, but generally suspect, reliability. Issues of the general quality of the available data in relation to choice of method are therefore of particular concern.

Effects of Varying Levels of Unreliability

Chance (random) variation, such as might result from attempts to gather information from incomplete, redundant, or confusing and contradictory base files,* could have quite different consequences given different methods of instrument development. As discussed in an earlier section, some methods (typically the more sophisticated) make greater "use" of the available data and their characteristics; hence, they may be expected to capitalize more on chance variation. This, of course, should result in differential shrinkage in validation. No such differential shrinkage was observed in the studies reported above.

This absence of differential shrinkage may be due, in part, to the careful and systematic attention paid to the reduction of the very large number of potential predictors to a smaller number for analysis, and, in part, to the nature of the data and of the problem. (All regression weights, for example, are small and all are nearly equal.) It could also be the case, however, that little in the way of random or chance variation is present in either the construction or the validation sample (although this is doubtful).

*It could well be the case, of course, that such variation may be systematic rather than random (particularly in the case of incomplete or missing files). It may be assumed, however, that there always will be a component of variable (random) error in any data elements coded from case files. That is, one may assume that any data element is comprised of a "time" value, plus a "constant error" (e.g., bias due to procedures or coders) plus a "variable error": $X = x_t + x_{ce} + x_{ve}$.

Nonetheless, a comparison of these devices given varying degrees of known (un)reliability may be of interest. The question may be posed relatively simply: Given the systematic introduction of known degrees of random error into the data base, do the different devices evince differential shrinkage in terms of predictive utility? [56]

Since the 1970 (construction) sample data are known to have reasonably high reliability (of one sort, at any rate), this sample was chosen for study. In the first phase, 10 percent random error was introduced into the predictor-item pool.* That is, with the probability of 0.10, any given data element may have been changed to some other coding category (given only that the code selected was one used for the item). Each model then was "validated" on the resulting "perturbed" data set.† The process was then repeated except that the probability of change (i.e., the level of added error introduced) was increased to 0.30.

Table 20 and Figure 9 display these results, along with the original construction and validation study findings. Not until we reach the highest "level" of perturbation (30 percent) do substantive conclusions change. Indeed, the striking result is the remarkable stability that all devices demonstrate (again with the exception of the Association Analysis). While differences are small, the Burgess device performs better given the severely perturbed data set.

Although all devices discussed may indeed result in the same degree of predictive utility, the different devices might still tend to classify the same individual differently. To investigate this issue, we examined the interrelations among the expected outcomes generated by various devices. Some methods (the clustering and the log-linear) result in "predictions" which are typological only. Hence, we based these comparisons on the expected probability level associated with membership in a given stratum (terminal cluster or cell for the clustering and log-linear models; grouped or collapsed probability levels for the Burgess and regression models). This was the comparative basis for each operational instrument (as opposed to the original equation or model developed).††

Table 21 presents the results of this investigation. With the exception of the Association Analysis (which performed substantially less powerfully in all analyses), the interrelations among devices are quite high. In no case (again, with the exception of the Association Analysis) is less than 50 percent of the variation in one instrument accounted for by another, and the more typical proportion of variance explained is on the order of 70-75 percent. Indeed, instruments developed using any of the three

*Since the original reliability of the criterion is unknown, it was not perturbed.

†A related issue concerns the effect of varying levels of reliability upon the replicability of the analyses (rather than on the validity of results). This question also may be of interest but it was not investigated in this study.

††Analyses were repeated using odds ratios and log-odds ratios instead of the simple probabilities. All results are substantively identical.

methods most commonly employed for such purposes (Burgess, Multiple Regression, and Predictive Attributes Analysis) are all very highly intercorrelated.*

Does It Matter?

The striking conclusion to which all analyses presented lead is that, given the types, measurement levels, and sophistication of available data and outcome criteria, no one method for developing operationally useful statistical decisionmaking aids provides an empirical advantage over the others considered. If this is indeed the case, then one might well ask, "Does it matter?"

Given the available evidence — from this study as well as from others cited earlier — the answer seems to be, "No." No clear-cut empirical advantage in prediction is provided by one or another method.

Empirical advantages, however, may not be the only—nor even the primary—advantages that may accrue given the use of different methods. For example, as mentioned before, some methods may be easier to implement simply because their procedures may be more readily understood, and hence utilized, by decisionmakers.

Again, we stress that the analyses we report here are based on data that may not be optimal for a comparison of methods. Until better predictor and outcome information is available, however, a reasonable conclusion would seem to be that, "It don't make no nevermind." [57]

*This could, of course, provide a practical and policy-relevant benefit, since it suggests that, by and large, each device would recommend a similar decision for the same person. Should this not prove the case, one could well imagine legal action alleging discriminatory treatment based on the properties of different statistical decisionmaking aids. We thank Professor Leslie T. Wilkins for suggesting this issue and the analysis described (Wilkins, L.T., personal communication, 1979).

Notes

[1]For reviews of the extensive parole prediction literature, see Mannheim, H. and Wilkins, L.T., *Prediction Methods in Relation to Borstal Training,* London: Her Majesty's Stationery Office, 1955; Gottfredson, D.M., "Assessment and Prediction Methods in Crime and Delinquency," *Task Force Report: Juvenile Delinquency and Youth Crime,* Washington, D.C.: United States Government Printing Office, 1967; Simon, H.F., *Prediction Methods in Criminology,* London: Her Majesty's Stationery Office, 1971.

[2]Gough, H.G., "Clinical versus Statistical Prediction in Psychology," *Psychology in the Making,* ed. L. Postman, New York: Knopf, 1962, emphasis in original.

[3]A similar view was voiced as early as 1941 by Paul Horst ("The Prediction of Personal Adjustment," New York: *Social Science Research Bulletin No. 48,* 1941), who suggested less quarreling and more borrowing. See also DeGroot, A.D., "Via Clinical to Statistical Prediction," invited address, Western Psychological Association, San Jose, April 1960.

[4]Meehl, P.E., *Clinical versus Statistical Prediction,* Minneapolis: University of Minnesota Press, 1954.

[5]Gough, H.G., *op. cit.,* 1962.

[6]Gottfredson, D.M., Wilkins, L.T., Hoffman, P.B., and Singer, S.M., *The Utilization of Experience in Parole Decision-Making: Summary Report,* Washington, D.C.,: United States Government Printing Office, November 1974 (see also Supplemental Reports 1 through 13, available from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22151); Simon, F.H., *op. cit.,* 1971.

[7]See, for example, Gottfredson, D.M., and Ballard, K.B., Jr., "Association Analysis, Predictive Attribute Analysis and Parole Behavior," paper presented at Western Psychological Association Meetings, Portland, Oregon, 1964; Simon, *op. cit.,* 1971; Wilbanks, W. and Hindelang, M., "The Comparative Efficiency of Three Predictive Methods." Appendix B in Gottfredson, D.M., Wilkins, L.T., and Hoffman, P.B., *Summarizing Experience for Parole Decisionmaking,* National Council on Crime and Delinquency Research Center, Davis, California, February 1972. Gottfredson et al., *op. cit.,* 1974; Solomon, H., "Parole Outcome: A Multidimensional Contingency Table Analysis," *Journal of Research in Crime and Delinquency,* 1976. Gottfredson, S.D., Gottfredson, D.M., and Wilkins, L.T., "A Comparison of Prediction Methods," Rutgers University, School of Criminal Justice, 1978. Van Alstyne, D.J., and Gottfredson, M.R., "A Multidimensional Contingency Table Analysis of Parole Outcome," *Journal of Research in Crime and Delinquency,* July 1978.

[8]Wilbanks, W. and Hindelang, M., *op. cit,* 1972.

[9]Lanyon, R.I. and Goodstein, L.D., *Personality Assessment,* New York: Wiley, 1971.

[10]Meehl, P.E. and Rosen, A., "Antecedent Probability and the Efficiency of Psychometric Signs, Patterns or Cutting Scores," *Psychological Bulletin, 52,* 1955. The statement is equally applicable to devices other than "tests."

[11]*Ibid.*

[12]*Ibid.*

[13]Fisher, J., "The Twisted Pear and the Prediction of Behavior," *Journal of Consulting Psychology, 23,* 1959; Cronbach, L.J., *Essentials of Psychological Testing,* New York: Harper Brothers, 1960.

[14]Ohlin, L.E. and Duncan, O.D., "The Efficiency of Prediction in Criminology," *American Journal of Sociology, 54,* 1949.

[15]Duncan, O.D. and Duncan, B., "A Methodological Analysis of Segregation Indexes," *American Sociological Review, 20,* 1955.

[16]Other indices are available, such as described by Richardson (Richardson, M.W., "Effectiveness of Selection Devices," *Handbook of Applied Psychology,* Vol. 1, eds. D.H. Fryer and E.R. Henry, New York: Rhinehard, 1950), but they often require information unavailable in parole prediction studies.

[17]Lancucki, L., and Tarling, R., "The Relationship Between Mean Cost Rating (MCR) and Kendall's Rank Correlation Coefficient," in Gottfredson, D.M., Wilkins, L.T., and Hoffman, P.B., *Guidelines for Parole and Sentencing Decisions: A Policy Control Method,* Boston: Lexington Books, 1978.

[18]Furgusson, D.M., Fifield, J.K., and Slater, S.W., "Signal Detectability Theory and the Evaluation of Prediction Tables," *Journal of Research in Crime and Delinquency, 14,* 1977. For a discussion of signal detection theory, see Green, D.M., and Swets, J.A., *Signal Detection Theory and Psychophysics,* New York: Wiley, 1966.

[19]Campbell, D.T. and Stanley, J.C., *Experimental and Quasi-Experimental Designs for Research,* Chicago: Rand McNally, 1963.

[20]Cronbach, L.J., *op.cit.,* 1960.

[21]Gottfredson, D.M., "Diagnosis, Classification, and Prediction," *Decision-Making in the Criminal Justice System: Reviews and Essays,* ed. D.M. Gottfredson, Washington, D.C.: United States Government Printing Office, 1975. See also Gottfredson, D.M., and Gottfredson, M.R., "Data for Criminal Justice Evaluation: Some Resources and Pitfalls" in Klein, M.W., and Teilman, K.S., eds., *Handbook of Criminal Justice Evaluation,* Beverly Hills: Sage, *in press.*

[22]Cureton, E.F., "Validity, Reliability, and Baloney," *Problems in Human Assessment,* eds. D.M. Jackson and S. Messick, New York: McGraw-Hill, 1967.

[23]Gottfredson, D.M., *op.cit.,* 1975.

[24]See, for example, O'Leary, V.O., Gottfredson, M.R., and Gelman, A., "Contemporary Sentencing Proposals," *Criminal Law Bulletin,* September-October 1975, pp. 555-586.

[25]For a further discussion and examples of the applications discussed, see Gottfredson, D.M., *op.cit.,* 1975; concerning inclusion of prediction instruments in decision guidelines models, see Gottfredson, D.M. *et al., op cit.,* 1978.

[26]Meehl, P.E., *op.cit.,* 1954.

[27]Gough, H.G., *op.cit.,* 1962.

[28]Mannheim, H. and Wilkins, L.T., *op. cit.,* 1955.

[29]Simon, F.H., *op. cit.,* 1971.

[30]DeSoto, C.B., "The Predilection for Single Orderings," *Journal of Abnormal and Social Psychology, 62,* 1961.

[31]See Porebski, O.R., "On the Interrelated Nature of the Multivariate Statistics Used in Discrimination Analysis," *British Journal of Mathematical and Statistical Psychology,* 1966, *19* (2), 197-214, p. 202.

[32]Cohen, J. and Cohen, P., *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences,* New York: Wiley, 1975.

[33]Dawes, R.M. and Corrigan, B., "Linear Models in Decision-Making," *Psychological Bulletin, 81,* 1974.

[34]Wainer, H., "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin, 83,* 1976.

[35]Gottfredson, D.M., Wilkins, L.T., Hoffman, P.B., and Singer, S.M., *op.cit.,* 1974; see especially Wilkins, L.T., *Inefficient Statistics,* Supplementary Report Six, and Wilkins, L.T., *Information Overload: Peace or War with the Computer,* Supplementary Report Eleven.

[36]Simon, F.H., *op.cit.,* 1971.

[37]Burgess, E.W., "Factors Determining Success or Failure on Parole," *The Workings of the Indeterminate Sentence Law and the Parole System in Illinois,* ed. A.A. Bruce, Springfield, Illinois: Illinois State Board of Parole, 1928.

[38]Wilkins, L.T. and MacNaughton-Smith, P., "New Prediction and Classification Methods in Criminology," *Journal of Research in Crime and Delinquency,* 1954, *1,* pp. 19-32.

[39]Johnson, S.C., "Hierarchical Clustering Schemes," *Psychometrika, 32,* 1967.

[40]Cormack, R.M., "A Review of Classification," *Journal of the Royal Statistical Society,* A, 1971.

[41]Cormack, R.M., ibid.; Lance, G.N. and Williams, W.T., "Computer Programs for Monothetic Classification," *The Computer Journal,* 1966; and Johnson, *op. cit.,* 1967, have reviewed specifications for several cluster schemes.

[42]Babst, D.V., Gottfredson, D.M., and Ballard, K.B., Jr., "Comparison of Multiple Regression and Configural Analysis Techniques for Developing Base Expectancy Tables," *Journal of Research in Crime and Delinquency, 5,* 1968.

[43]Cormack, R.M., *op.cit.,* 1971.

[44]Williams, W.T. and Lambert, J.M., "Multivariate Methods in Plant Ecology I, Association-Analysis in Plant Communities," *J.Ecol., 47,* 1959; Williams, W.T. and Lambert, J.M., "Multivariate Methods in Plant Ecology II, The Use of an Electronic Digital Computer for Association-Analysis," *J. Ecol., 49,* 1961; Gottfredson, D.M., Ballard, K.B., Jr., and Lane, L., *Association Analysis in a Prison Sample and Prediction of Parole Performance,* Vacaville, California: Institute for the Study of Crime and Delinquency, November 1963.

[45]Ballard, K.B., Jr. and Gottfredson, D.M., *Predictive Attribute Analysis and Prediction of Parole Performance,* Vacaville, California: Institute for the Study of Crime and Delinquency, December 1963.

[46]Wilkins, L.T. and MacNaughton-Smith, P. *op.cit,* 1954.

[47]*Ibid.*

[48]For a general consideration of the technique, see Goodman, L.A., "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications," *Journal of the American Statistical Association,* 1970, *65,* 226-265, Goodman, L.A., "The Analysis of Multi-dimensional Contingency Tables: Step-wise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications, *Technometrics,* 1971, *13,* 33-61. For an example of the application of the method to the assessment of risk, see Solomon, H., "Parole Outcome: A Multidimensional Contingency Table Analysis," *Journal of Research in Crime and Delinquency,* 1976, *13,* 107-126. In essence, logit analysis is an extension of a somewhat earlier model (probit analysis; see Ashton, W.D., *The Logit Transformation,* New York: Hofner, 1972) designed to overcome limitations of standard regression techniques (see Nerlov, M. and Press, S.J., *Univariate and Multivariate Log-Linear and Logistic Models,* Santa Monica, California: The Rand Corp., 1973).

[49]Gottfredson, D. M., Wilkins, L. T., Hoffman, P. B., and Singer, S. M., *op.cit.,* 1974.

[50]Gottfredson, D.M., Wilkins, L.T., Hoffman, P.B., and Singer, S.M., *op.cit.,* 1974.

[51]Wilkins, L.T. and MacNaughton-Smith, P., *op cit.,* 1964.

[52]Johnson, S.C., *op cit.,* 1967.

[53]Simon, F.H., *op cit.,* 1971.

[54]See Magidson, J., "An illustrative comparison of Goodman's approach to Logit Analysis with dummy variable regression analysis" chapter 2 in Goodman, L.A., *Analyzing Qualitative/Categorical Data,* Cambridge, Mass.: Abt Books, 1978.

[55]Guilford, J.P., *Fundamental Statistics in Psychology and Education,* New York: McGraw-Hill, 1965.

[56]Wilbanks, W. and Hindelang, M., *op cit.,* 1972.

[57]Wainer, *op cit.,* 1976.

Table A-1
CORRELATIONS OF PRESENT OFFENSE VARIABLES

	1	2	3	4	5	6	7	8
1. **Outcome**[1] (N = 2,493)	—	-0.10	0.10	-0.15	-0.15	-0.20	-0.19	-0.10
2. **How Committed:** All Probation, Parole, and Mandatory Release Violations vs. All Other Commitments (N = 2,497)		—	-1.00	0.10	0.10	0.16	0.15	0.13
3. **Type of Admission:** New Court Commitment vs. All Other (N = 2,497)			—	-0.10	-0.10	-0.16	-0.15	-0.13
4. **Burglary, Larceny, Theft, and Fraud Offenses, and "All Other"** vs. All Other Offenses (N = 2,496)				—	0.82	0.42	0.68	-0.03
5. **Person and Property Offenses**[2] vs. All Other Offenses (N = 2,496)					—	0.24	0.56	0.02
6. **Vehicle Theft, Forgery, Fraud, Larceny by Check, Heroin** vs. All Other Offenses (N = 2,496)						—	0.82	0.12
7. **Vehicle Theft, Forgery, Fraud, Larceny by Check** vs. All Other Offenses (N = 2,496)							—	0.18
8. **Dollar Value of Crime:** Up to $499 or "Unknown" vs. $500 and above (N = 2,497)								—

[1]Criterion A; Table 2.
[2]Includes homicide, manslaughter, all robbery, assault, all burglary, all theft.

Table A-2
CORRELATIONS OF CRIMINAL HISTORY VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Outcome[1] (N = 2,493)	—	-0.18	-0.17	-0.18	-0.23	0.16	0.15	0.10	0.20	0.18	0.12	0.10	0.19	0.13	0.24	0.17	-0.12
2. Age at First Arrest: 14 and under; 15-17; 18-21; 22 and over (N = 2,496)		—	0.83	0.73	0.28	-0.08	-0.33	-0.29	-0.27	-0.23	-0.32	-0.17	-0.29	0.06	-0.31	-0.22	0.36
3. Age at First Conviction: 15 and under; 16-18; 19-22; 23 and over (N = 2,496)			—	0.81	0.31	-0.08	-0.37	-0.32	-0.30	-0.24	-0.34	-0.17	-0.30	0.07	-0.32	-0.28	0.35
4. Age at First Commitment: 17 and under; 18-20; 21-25, 26 and over (N = 2,495)				—	0.33	-0.13	-0.26	-0.16	-0.36	-0.25	-0.32	-0.16	-0.32	0.04	-0.32	-0.32	0.33
5. Longest Time Free Since First Commitment: 6 Months or less; 7-18 Months; 19-36 Months; 37-60 Months; More than 60 Months (N = 2,497)					—	-0.24	-0.37	-0.15	-0.41	-0.39	-0.21	-0.16	-0.30	-0.11	-0.37	0.27	0.20
6. Number of Prior Convictions: None; One; 2-3; 4 or More (N = 2,497)						—	0.23	0.03	0.69	0.33	0.27	0.31	0.30	0.35	0.53	0.67	-0.14
7. Number of Prior Sentences: None; 1-2; 3 or More (N = 2,495)							—	0.48	0.52	0.30	0.30	0.35	0.22	0.19	0.50	0.37	-0.17
8. Sentences with Probation: None; One; 2 or More (N = 2,495)								—	0.18	0.29	0.24	0.22	0.17	0.11	0.36	0.13	-0.19
9. Number of Prior Incarcerations: None; One; 2 or More (N = 2,495)									—	0.40	0.34	0.46	0.39	0.31	0.64	0.63	-0.20
10. Probation or Parole Revocation: None vs. Any Revocation (N = 2,490)										—	0.17	0.17	0.25	0.17	0.34	0.32	-0.19
11. Prior Convictions for Burglary: No Prior Convictions for Burglary (N = 2,497)											—	0.15	0.14	-0.02	0.47	0.34	-0.29
12. Prior Convictions for Larceny: No Prior Convictions for Larceny (N = 2,497)												—	0.06	0.10	0.54	0.25	-0.19
13. Prior Convictions for Auto Theft: No Prior Convictions for Auto Theft (N = 2,497)													—	0.05	0.53	0.30	-0.28
14. Prior Convictions for Forgery, Fraud, or Larceny by Check: No Prior Convictions for Forgery, Fraud, or Larceny by Check (N = 2,497)														—	0.44	0.22	-0.10
15. Total Convictions for Property Offenses: 0-1 Arrests; 2 Arrests; 3 Arrests; 4 or More Arrests (N = 2,497)															—	0.52	-0.37
16. Longest Time Served on Any Commitment: 0-6 Months vs. More than 6 Months																—	-0.22
17. Reason for 1st Arrest: Burglary, Check Offenses, Forgery, Theft, Delinquent Child vs. All Others																	—

[1]Criterion A; Table 2.

Table A-3
CORRELATIONS OF CRIMINAL HISTORY VARIABLES (Dichotomized)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Outcome[1] (N = 2,493)	—	-0.17	-0.16	-0.17	0.23	0.13	0.14	0.09	0.17	0.18	0.12	0.10	0.17	0.12	0.21	0.17	-0.12
2. Age at First Arrest: Less than 18 years vs. 18 years or older		—	0.84	0.65	-0.24	-0.08	-0.32	-0.25	-0.26	-0.20	-0.25	-0.18	-0.31	-0.01	-0.28	-0.22	0.36
3. Age at First Conviction: Less than 18 years vs. 18 years or older			—	0.77	-0.25	-0.08	-0.35	-0.28	-0.27	-0.21	-0.29	-0.18	-0.32	0.01	-0.29	-0.24	0.33
4. Age at First Commitment: Less than 18 years vs. 18 years or older				—	0.25	0.11	-0.25	0.15	-0.34	-0.22	-0.30	-0.15	-0.32	0.02	-0.29	-0.31	0.31
5. Longest Time Free Since First Commitment: None or 60 or More Months vs. All Other (N = 2,497)					—	0.16	0.24	0.11	0.42	0.31	0.17	0.11	0.28	0.09	0.29	0.27	-0.16
6. Number of Prior Convictions: None vs. All Other (N = 2,497)						—	0.21	0.04	0.56	0.30	0.26	0.20	0.16	0.26	0.38	-0.27	-0.13
7. Number of Prior Sentences: None vs. All Other (N = 2,495)							—	0.42	0.48	0.26	0.24	0.30	0.22	0.18	0.43	0.32	-0.20
8. Sentences with Probation: None vs. All Other (N = 2,495)								—	0.19	0.29	0.20	0.21	0.17	0.13	0.35	0.14	-0.18
9. Prior Incarcerations: None vs. All Other (N = 2,495)									—	0.32	0.29	0.27	0.27	0.24	0.47	0.63	-0.20
10. Probation or Parole Revocation: None vs. All Other (N = 2,490)										—	0.17	0.16	0.21	0.17	0.29	0.32	-0.19
11. Prior Convictions for Burglary: None vs. All Other (N = 2,497)											—	0.14	0.17	0.04	0.43	0.34	-0.29
12. Prior Convictions for Larceny: None vs. All Other (N = 2,497)												—	0.05	0.16	0.44	0.23	-0.22
13. Prior Convictions for Auto Theft: None vs. All Other (N = 2,497)													—	0.37	0.46	0.26	-0.31
14. Prior Convictions for Forgery, Fraud, or Larceny by Check: None vs. All Other (N = 2,497)														—	0.39	0.23	-0.12
15. Total Convictions for Property Offenses: 0-1 Arrests vs. All Other (N = 2,497)															—	0.46	-0.38
16. Longest Time Served on Any Commitment: 0-6 Months vs. More than 6 Months																—	-0.22
17. Reason for 1st Arrest: Burglary, Check Offenses, Forgery, Theft, Delinquent Child vs. All Others																	—

[1]Criterion A; Table 2

Table A-4
CORRELATIONS OF SOCIAL HISTORY VARIABLES

	1	2	3	4	5	6	7
1. **Outcome**[1] (N = 2,493)	—	-0.08	0.14	0.14	0.09	0.13	-0.14
2. **Highest Grade Completed:** 0-11 vs. All Other (N = 2,497)		—	-0.04	-0.05	-0.04	-0.17	0.04
3. **Marital Status at Admission:** Married, or Common-Law Spouse vs. All Other (N = 2,483)			—	0.78	0.36	0.11	-0.18
4. **Living Arrangement Before Commitment:** Wife and/or Children vs. All Other (N = 2,497)				—	0.08	0.14	-0.24
5. **Use of Synthetic and/or Natural Opiates:** Use vs. Non-Use (N = 2,497)					—	0.12	-0.06
6. **Employment in Last 2 Years of Civilian Life:** Unemployable or Employed[2] vs. Unemployed (N = 2,497)						—	-0.25
7. **Longest Job in Free Community:** No Job Through 4 Years of Employment, and Unknown vs. All Other (N = 2,497)							—

[1]Criterion A; Table 2.

[2]Employed more than 25% of the time, or student, or unemployable 75% of the time.

Table A-5
CORRELATIONS OF SOCIAL HISTORY VARIABLES (Dichotomized)

	1	2	3	4	5	6	7
1. **Outcome**[1] (N = 2,493)	—	-0.07	0.14	0.14	0.09	0.13	-0.16
2. **Highest Grade Completed:** No Schooling Through Ph.D. (N = 2,497)		—	-0.14	-0.04	0.00	-0.19	0.07
3. **Marital Status at Admission:** Married, or Common-Law Spouse vs. All Other (N = 2,483)			—	0.78	0.04	0.11	-0.22
4. **Living Arrangement Before Commitment:** Wife and/or Children vs. All Other (N = 2,497)				—	0.08	0.14	-0.26
5. **Use of Synthetic and/or Natural Opiates:** Use vs. Non-Use (N = 2,497)					—	0.12	-0.02
6. **Employment in Last 2 Years of Civilian Life:** Unemployable or Employed[2] vs. Unemployed (N = 2,497)						—	-0.39
7. **Longest Job in Free Community:** Less Than One Year or Unknown vs. From One to Four Years vs. More Than Four Years (N = 2,497)							—

[1]Criterion A; Table 2.

[2]Employed more than 25% of the time, or student, or unemployable 75% of the time.

Table A-6
CORRELATIONS OF INSTITUTIONAL ADJUSTMENT VARIABLES

	1	2	3	4
1. **Outcome**[1] (N = 2,493)	—	0.13	0.07	0.12
2. **Escape History:** None vs. One or More Escapes (N = 2,497)		—	0.16	0.18
3. **Latest Custody Classification:** Maximum or Close; Medium; Minimum or Work-Release (N = 2,497)			—	0.10
4. **Prison Punishment:** None vs. Punishment Received (N = 2,497)				—

[1]Criterion A; Table 2.

Table A-7
CORRELATIONS OF INSTITUTIONAL ADJUSTMENT VARIABLES (Dichotomized)

	1	2	3	4
1. **Outcome**[1] (N = 2,493)	—	0.13	0.12	0.09
2. **Escape History:** No Escapes vs. One or More (N = 2,497)		—	0.18	0.21
3. **Prison Punishment:** None vs. Punishment Received (N = 2,497)			—	0.14
4. **Latest Custody Classification:** Unknown, Minimum, or Work-Release vs. Maximum, Close, or Medium (N = 2,497)				—

[1]Criterion A; Table 2.

Table A-8
CORRELATIONS OF PRESENT OFFENSE VARIABLES (Dichotomized) WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
How Committed: All Probation, Parole, and Mandatory Release Violations vs. All Other	-0.10	-0.12	-0.12	-0.12
Type of Admission: New Court Commitment vs. All Other	0.10	0.12	0.12	0.12
All Burglary, Larceny, Theft, and Fraud Offenses vs. All Other Offenses	-0.15	-0.16	-0.17	-0.17
Dollar Value of Crime: Up to $499 or "Unknown" vs. All Other	-0.10	-0.12	-0.10	-0.12

CONTINUED

1 OF 2

Table A-9
CORRELATIONS OF CRIMINAL HISTORY VARIABLES WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Age at First Arrest: 14 or under; 15-17; 18-21; 22 or over	-0.17	-0.18	-0.21	-0.19
Age at First Conviction: 15 or under; 16-18; 19-22; 23 or over	-0.17	-0.18	-0.20	-0.19
Age at First Commitment: 17 or under; 18-20; 21-25; 26 or over	-0.18	-0.19	-0.21	-0.19
Longest Time Free Since First Commitment: 6 mos or less; 7-18 mos; 19-36 mos; 37-60 mos; over 60 mos (includes Code of '0')	-0.23	-0.26	-0.29	-0.27
Number of Prior Convictions: None; 1; 2-3; 4 or More	0.15	0.14	0.15	0.14
Number of Prior Sentences: None; 1-2; 3 or More	0.15	0.18	0.21	0.19
Number of Prior Sentences with Probation: None; 1; 2 or More	0.10	0.12	0.11	0.13
Number of Prior Incarcerations: None; 1; 2 or More	0.20	0.21	0.25	0.22
Number of Prior Parole/Probation Revocations: None vs. Any	0.18	0.19	0.19	0.18
Number of Prior Convictions for Burglary: No Prior Convictions for Burglary	0.12	0.14	0.15	0.14
Number of Prior Convictions for Larceny: No Prior Convictions for Larceny	0.10	0.10	0.15	0.11
Number of Prior Convictions for Auto Theft: No Prior Convictions for Auto Theft	0.19	0.20	0.19	0.20
Number of Prior Convictions for Forgery, Fraud, or Larceny by Check: No Prior Convictions for Forgery, Fraud, or Larceny by Check	0.13	0.11	0.11	0.11
Total Number of Prior Convictions for Property Offenses: 0-1; 2; 3; 4 or More	0.23	0.24	0.26	0.26
Reason for First Arrest: Burglary, Check Offenses, Forgery, Theft, Delinquent Child vs. All Others	-0.12	-0.11	-0.11	-0.12
Longest Time Served on Any Commitment: 0-6 mos vs. More than 6 mos	0.17	0.18	1.19	0.20

Table A-10
CORRELATIONS OF CRIMINAL HISTORY VARIABLES (Dichotomized) WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Age at First Arrest: 18 or under; over 18	-0.17	-0.18	-0.21	-0.19
Age at First Conviction: 18 or under; over 18	-0.16	-0.17	-0.19	-0.17
Age at First Commitment: 18 or under; over 18	-0.17	-0.17	-0.18	-0.17
Longest Time Free Since First Commitment: Code '0' and more than 60 mos vs. Less than 60 mos	0.23	0.26	0.27	0.26
Number of Prior Convictions: None vs. Any	0.13	0.13	0.14	0.13
Number of Prior Sentences: None vs. Any	0.14	0.16	0.19	0.17
Number of Prior Sentences with Probation: None vs. Any	0.09	0.11	0.10	0.12
Number of Prior Incarcerations: None vs. Any	0.17	0.19	0.23	0.20
Number of Prior Parole/Probation Revocations: None vs. Any	0.18	0.19	0.19	0.20
Number of Prior Convictions for Burglary: None vs. Any	0.12	0.14	0.15	0.14
Number of Prior Convictions for Larceny: None vs. Any	0.10	0.11	0.14	0.12
Number of Prior Convictions for Auto Theft: None vs. Any	0.17	0.19	0.21	0.20
Number of Prior Convictions for Forgery, Fraud, Larceny by Check: None vs. Any	0.12	0.11	0.12	0.11
Reason for First Arrest: Burglary, Check Offenses, Forgery, Theft, Delinquent Child vs. All Others	-0.12	-0.11	-0.12	-0.11
Longest Time Served on Any Commitment: 0-6 mos vs. More Than 6 mos	0.17	0.18	.20	0.19

Table A-11
CORRELATIONS OF SOCIAL HISTORY VARIABLES WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Highest Grade Completed: No Schooling Through Ph.D.	-0.07	-0.09	-0.10	-0.10
Marital Status at Admission: Married, or Common-Law Spouse vs. All Other	0.14	0.14	0.15	0.15
Living Arrangement Before Commitment: Wife and/or Children vs. All Other	0.14	0.16	0.18	0.17
Use of Synthetic and/or Natural Opiates: Use vs. Non-Use	0.09	0.08	0.13	0.09
Employment in Last 2 Years of Civilian Life: Unemployable or Employed vs. Unemployed	0.13	0.14	0.15	0.15
Longest Job in Free Community: Less Than 1 Year or Unknown vs. From 1 to 4 Years vs. More Than 4 Years	-0.16	-0.16	-0.18	-0.16

Table A-12
CORRELATIONS OF SOCIAL HISTORY VARIABLES (Dichotomized) WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Highest Grade Completed: 0-11 vs. All Other	-0.07	-0.09	-0.10	-0.10
Marital Status at Admission: Married, or Common-Law Spouse vs. All Other	0.14	0.14	0.15	0.15
Living Arrangement Before Commitment: Wife and/or Children vs. All Other	0.14	0.16	0.18	0.17
Use of Synthetic and/or Natural Opiates: Use vs. Non-Use	0.09	0.08	0.13	0.09
Employment in Last 2 Years of Civilian Life: Unemployable or Employed vs. Unemployed	0.13	0.14	0.15	0.15
Longest Job in Free Community: No Job Through 4 Years of Employment, and Unknown vs. All Other	-0.14	-0.15	-0.18	-0.15

Table A-13
CORRELATIONS OF INSTITUTIONAL ADJUSTMENT VARIABLES WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Escape History: None vs. One or More Escapes	0.13	0.16	0.17	0.17
Latest Custody Classification: Maximum or Close; Medium; Minimum or Work-Release	0.07	0.09	0.08	0.09
Prison Punishment: None vs. Punishment Received	0.12	0.14	0.16	0.15

Table A-14
CORRELATIONS OF INSTITUTIONAL ADJUSTMENT VARIABLES (Dichotomized) WITH OUTCOMES

	Outcome A (N = 2,493)	Outcome B (N = 2,493)	Outcome C (N = 2,382)	Outcome D (N = 2,382)
Escape History: No Escapes vs. One or More	0.13	0.16	0.17	0.17
Prison Punishment: None vs. Punishment Received	0.12	0.14	0.16	0.15
Latest Custody Classification: Unknown, Minimum, or Work-Release vs. Maximum, Close, or Medium	0.09	0.10	0.11	0.11

Table A-15
NINETEEN-ITEM BURGESS SCALE/OUTCOME DISTRIBUTION EQUAL-STEP COLLAPSE

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
0-3	305	37.4
4-7	740	54.9
8-11	656	64.3
12-15	491	82.5
16-19	190	96.3
TOTAL GROUP	2,382	64.2

Eta = 0.343 r_{pb} = 0.340 MCR = 0.394
Eta² = 0.118 r_{pb}^2 = 0.116

Table A-16
NINETEEN-ITEM BURGESS SCALE/OUTCOME DISTRIBUTION PENTILES

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
0-4	477	44.0
5-7	568	54.6
8-10	504	60.7
11-13	410	79.3
14-19	423	89.6
TOTAL GROUP	2,382	64.2

Eta = 0.336 r_{pb} = 0.332 MCR = 0.389
Eta² = 0.113 r_{pb}^2 = 0.110

Table A-17
NINETEEN-ITEM BURGESS SCALE/OUTCOME DISTRIBUTION HALF-STANDARD DEVIATION STEPS

(Construction Sample)

Burgess Score	N	Percent Favorable Outcome
0-2	164	32.9
3-4	313	49.8
5-6	357	50.7
7-8	393	58.3
9-11	474	67.9
12-13	258	81.0
14-15	233	84.1
16+	190	96.3
TOTAL GROUP	2,382	64.2

Eta = 0.350 r_{pb} = 0.345 MCR = 0.408
Eta² = 0.123 r_{pb}^2 = 0.119

Table A-18
MULTIPLE REGRESSION OF PRESENT OFFENSE VARIABLES ON OUTCOME[1]

Variable[2]	Unstandardized Weight	Standardized Weight	R	R^2
Commitment Offense	-0.1519	-0.1547	0.165	0.027
Type of Admission	0.1209	0.1062	0.196	0.039
Constant	1.5411			

[1]See footnote 1, Table 10.
[2]See footnote 2, Table 10.

Table A-19
MULTIPLE REGRESSION OF CRIMINAL HISTORY VARIABLES ON OUTCOME[1]

Variable[2]	Unstandardized Weight	Standardized Weight	R	R^2
Longest Time Free	-0.0464	-0.1588	0.269	0.073
Crime Group	0.0372	0.0904	0.318	0.101
Age at First Arrest	-0.0969	-0.0939	0.330	0.109
Number of Prior Parole/ Probation Revocations	0.0290	0.0595	0.335	0.112
Number of Prior Commitments for Auto Theft	0.0281	0.0709	0.338	0.114
Number of Prior Commitments for Forgery	0.0197	0.0433	0.339	0.115
Reason for First Arrest	0.0334	0.0348	0.340	0.115
Number of Prior Commitments for Burglary	0.0376	0.0353	0.341	0.117
Constant	1.5332			

[1]See footnote 1, Table 10.
[2]See footnote 2, Table 10.

Table A-20
MULTIPLE REGRESSION OF SOCIAL HISTORY VARIABLES ON OUTCOME[1]

Variable[2]	Unstandardized Weight	Standardized Weight	R	R^2
Planned Living Arrangement	0.1038	0.0859	0.166	0.028
Employment	0.0685	0.0715	0.211	0.044
Longest Job in Free Community	-0.0616	-0.0961	0.227	0.052
Known Use of Synthetic and/or Natural Opiates	0.0440	0.0684	0.237	0.056
Highest Grade Claimed	-0.0121	-0.0708	0.247	0.061
Living Arrangement	0.0766	0.0672	0.252	0.064
Constant	1.1687			

[1]See footnote 1, Table 10.
[2]See footnote 2, Table 10.

Table A-21
MULTIPLE REGRESSION OF INSTITUTIONAL ADJUSTMENT VARIABLES ON OUTCOME[1]

Variable[2]	Unstandardized Weight	Standardized Weight	R	R^2
Escape	0.1639	0.1369	0.167	0.028
Prison Punishment	0.1325	0.1249	0.210	0.044
Custody Classification	0.0334	0.0539	0.217	0.047
Constant	1.2385			

[1]See footnote 1, Table 10.
[2]See footnote 2, Table 10.

Table A-22
SELECTED "REDUCED MAIN EFFECTS" MODELS FROM THE 2^9 CONTINGENCY TABLE

H_i	Descriptions	χ^2 (d.f.)	$\chi^2_1 - \chi^2_i$ (p)	R^2_i	$\chi^2_i - \chi^2_2$ (p)	$r^2_{2.i}$
H_{11}	H_2, except PLIVARR and AGECON	191.01 (249)	298.47 (<0.005)	0.61	8.72 (<0.01)	0.05
H_{12}	H_2, except PLIVARR and MARITAL	200.76 (249)	288.72 (<0.005)	0.59	18.47 (<0.005)	0.09
H_{13}	H_2, except PLIVARR and ESCAPE	195.94 (249)	293.54 (<0.005)	0.60	13.65 (<0.005)	0.07
H_{14}	H_2, except AGECON and MARITAL	193.53 (249)	295.95 (<0.005)	0.60	11.24 (<0.005)	0.06
H_{15}	H_2, except AGECON and ESCAPE	200.07 (249)	289.41 (<0.005)	0.59	17.78 (<0.005)	0.09
H_{16}	H_2, except MARITAL and ESCAPE	198.37 (249)	291.11 (<0.005)	0.59	16.08 (<0.005)	0.08
H_{17}	H_2, except PLIVARR, AGECON, and MARITAL	206.76 (250)	282.72 (<0.005)	0.58	24.27 (<0.005)	0.12
H_{18}	H_2, except PLIVARR, AGECON, and ESCAPE	203.44 (250)	286.04 (<0.005)	0.58	21.15 (<0.005)	0.10
H_{19}	H_2, except AGECON, ESCAPE, and MARITAL	205.86 (250)	283.62 (<0.005)	0.58	23.57 (<0.005)	0.11
H_{20}	H_2, except PLIVARR, MARITAL, and ESCAPE	209.99 (250)	279.49 (<0.005)	0.58	27.70 (<0.005)	0.13
H_{21}	H_2, except PLIVARR, AGECON, MARITAL, and ESCAPE	218.19 (251)	271.29 (<0.005)	0.55	35.90 (<0.005)	0.16

END