In collaboration with
... States Board
of Parole

..., Chairman
William ... Howard, Jr.
... A. ...

This is a working draft of a report and is not for publication or citation. It has not been reviewed by the United States Board of Parole.

This is one of a series of reports by the Parole Decision-Making project. The overall aim of the study is the development and demonstration of model programs for provision of information to paroling authorities in order to improve parole decisions by an increased utilization of experience in these decisions. The program, which is being conducted in collaboration with the United States Board of Parole, is supported by a grant from the National Institute of Law Enforcement and Criminal Justice of the Law Enforcement Assistance Administration and is administered by the National Council on Crime and Delinquency Research Center.

Advisory groups include the National Probation and Parole Institutes' Advisory Committee (with representation from the Association of Paroling Authorities, the Interstate Probation and Parole Compact Administrators' Association, the United States Board of Parole, the Probation Division of the Administrative Office of the United States Courts, and the Advisory Council on Parole of the National Council on Crime and Delinquency) and a Scientific Advisory Committee selected by LEAA, the U.S. Board of Parole, and project staff.

The opinions expressed are those of the authors and do not necessarily reflect the views or endorsement of the above agencies or groups.

United States Board of Parole

George J. Reed, Chairman
William E. Amos
Curtis C. Crawford
William F. Howland, Jr.
Gerald E. Murch
Maurice H. Sigler
Paula A. Tennant
William T. Woodard, Jr.

Former Board of Parole Members

Walter Dunbar
Zeigel W. Neff
Charlotte Paul Reese

National Probation and Parole Institutes Advisory Committee

Charles P. Chew, Chairman
George F. Denton
Charles H. Lawson
Vincent O'Leary
Merrill E. Smith
George J. Reed
John A. Wallace

Scientific Advisory Committee

Herbert Solomon, Chairman
T. Conway Esselstyn
C. Ray Jeffery
Malcolm W. Klein
Charles Newman
Stanton Wheeler

Project Staff

Don M. Gottfredson, Codirector
Leslie T. Wilkins, Codirector
Peter B. Hoffman, Research Associate
Susan M. Singer, Project Coordinator
Lucille K. DeGostin, Research Assistant
LaVerne L. Holmes, Research Assistant
E. Cordelia Mapp, Research Assistant
Annie C. Miller, Research Assistant
Carol Ann Riccio, Research Assistant
Anne M. Porowski, Secretary

Former Project Staff

Barbara J. Allen
Barbara A. Bullock
Yvonne D. Colbert
Gabrielle Pagin Ficklin
Anne L. Hatfield
C. Ann Sadowsky

NCCD Research Center Staff Support

Marilyn M. Turner, Secretary to the Director
Max S. Zeigler, Data Processing Supervisor
Guy E. Pasela, Research Assistant
Judith Ingram, Research Assistant
Kathryn E. Berger, Clerk-Typist

SUMMARY

Statisticians have devised a variety of procedures for combining information (such as items concerning offenders taken from case files) in order to use them efficiently in predicting later behavior or administrative action (such as parole violation). The concept of efficiency can take a variety of meanings, but one meaning relates to the question of whether or not all the information is needed or contributes usefully to the accuracy or validity of the prediction.

Many items "overlap" with one another; that is, they are correlated among themselves. For example, auto thieves tend to be younger than offenders in general; persons with more prior convictions tend to have more prior arrests and sentences; and those with prior parole violations necessarily have had prior prison terms. Statisticians, therefore, have invented procedures which take such overlapping into account. When this is done it typically is found that only a few items, appropriately weighted, may be expected to do the work--in prediction--of a much larger number.

From various studies in correctional systems, however, it now appears that less sophisticated methods of combining the information--such as simply adding favorable items together without weighting--may end up, in practice, as better than the more sophisticated techniques. This curious result suggests not that the statistical theory is wrong but that the nature of the data does not satisfy the assumptions which are made in statistical theory.

An implication--thought to be extremely important for both research and practice--is that major advances in both must await the development of better quality data.

Every individual makes predictions of the likely behavior of those other individuals with whom he comes into contact. Without such predictions life would be impossible. Some persons may be more inclined than others to "do their own thing"--that is to say, they may be more eccentric--but, nonetheless, much common behavior is predictable within varying limits. In ordinary life we use our general experience to make statements about the likely behavior of others. In business we may use statistical tables or actuarial estimates of probabilities based on certain mathematical assumptions. Some of the methods which have a basis in the calculus of probabilities have been worked out for certain behaviors of offenders. It is in relation to these methods that some peculiar results have been recently observed. It seems that the reasons for these results may be traced back to factors in the operational aspects of the criminal justice system.

Let us try to discuss these rather indigestible statistical problems with some light-hearted examples. Suppose you had a "computer date." You had asked to be put in touch with a girl (we will not specify age) of below average height. You might think it wiser to specify "below average height" rather than "below average weight," since you hope to be friendly when you meet! You will, of course, realize that height and weight tend to go together (positively correlated), and although you may be unlucky and get a girl who is both short and heavy, this is a risk you may have decided to take. Whether this is a wise decision or not we may, at this stage, leave to the imagination. (How much time are you prepared to waste, either in obtaining more data or in meeting dates who turn out unsuitable?)

But to continue our specification, which we are attributing to you, let us say that you add "dark, black hair; straight, small nose; mild manners." Let us suppose that the computer dating service says that it has found a girl who fits this specification.

Now consider that you are to meet her at, either, Tokyo or Stockholm airport! You will appreciate that at Tokyo almost every girl (having left age unspecified) who steps from an aircraft has dark, black hair, and so on; and you would have great difficulty in finding your friend without some other details. Alternatively, suppose you prefer blondes, tall, with fine figures; you would have no difficulty in identifying your friend at Tokyo, whereas in Stockholm the situation would be very different.

The point of this is, of course, to make it obvious that information in one setting is useless as a means of discrimination, while in another setting the same information is quite useful and may well isolate a particular individual from the mass of others. In this example we have used Tokyo and Stockholm locations, but the same thing would apply if the location were in the form of another item of information. It is a feature of the addition of items of information that when we have any one item--say, the first item whatever it might be--we are usually able to make a better than chance guess as to what the second item might be. In terms of our girls, the specification of "black hair" makes it almost unnecessary to specify "dark eyes"; and, similarly, the specification of "blonde" makes it almost certain that we shall meet a blue-eyed girl: if we have strong feelings about the eye color and we object to blue eyes while preferring blondes, we are going to be a little difficult to satisfy!

We see that, given one piece of information, another piece of information <u>in addition</u> to that we already possess has a different value from that which it has when it stands alone. This is a case of one and one not making two, except in a few situations where the first and second items of information are uncorrelated. There is, however, one further point which we may identify in our example--namely, that the value of the information can be tested only by reference to some utility. It was assumed that we wanted to identify the girl and to meet her at an airport. If we had had other intentions, equally honorable, of course, the value of the information would have been differently assessed. Thus we can claim that whether a particular item of data is or is not "information" depends upon how well it assists us in arriving at a rational decision. Or, more generally, as mathematicians are apt to say, information is that which reduces uncertainty. An item--or an addition of an item--which does not reduce, or reduce further, our uncertainty is not information. What may be information regarding the likelihood of recidivism may or may not be information about amenability to a form of training. Whatever may reduce your uncertainty with respect to the identification of the girl may or may not reduce your uncertainty as to whether you might marry her at some time. Thus, there is no general "information" in this meaning of the term. To put the case within our own particular framework, it may be information for the policeman concerned with identification that a suspect has "blue eyes," but it is not information for the judge who is considering sentence. Now, clearly, we can assess our uncertainty only about one thing at a time; and, hence, we can only assess the power of information with respect to one thing at a time.

We now have two requirements which we can state about information when we are concerned with more than one item, singly, and with respect only to one external activity, namely, (1) to qualify as "information" data must reduce uncertainty and (2) the reduction of uncertainty can relate only to one factor at a time.

Now it may seem obvious that so long as the information we consider is "relevant to the decision," the more of it we can consider, the better should be the decision. But it is not so simple as that. Any item of information (in the lay meaning of the term) may be highly relevant when it stands on its own--that is to say, it is correlated with the criterion. But the same information item may not remain significant when taken into consideration with other items, because anything we may have learned about the criterion may already have been covered by the item or items already considered.

From the argument we have stated so far, we may have proved our previous claim--that, when adding items of information, one and one do not make two. If this is so, then we need some other convention for addition which will assist us in the use of cumulative information, because we want to make the best use of all that we might know. Statisticians have developed methods for this form of addition.

It is possible to examine a large body of data and find the one piece of information which, on its own, is the most useful in predicting a particular criterion. This would be that item which was most highly correlated with the criterion. Clearly we can select only one criterion at a time, because the item which is most highly correlated with one criterion may not be that which is most highly correlated with another criterion. When we have identified the most powerful item of information, we can search the field of information for another item which, given the first item, is then most highly correlated with the criterion. It is, of course, necessary to find a means for taking out of the reckoning the power of the first item before we add the second or even attempt to assess its contribution to the prediction of the criterion. This is usually termed the problem of "overlap." If two items of information are highly correlated with each other, then, when we have taken the first into consideration, the second will have lost much of its power because it is telling us "nothing new." In the same way we can go along searching for a third item which, in the presence of the previous two items, adds something new to our ability to predict the criterion. It is not usually long before it becomes very difficult to find items which, in the presence of those already included, add anything new to the prediction. The process described is, of course, known as "step-wise regression."

Step-wise regression is one of a family of similar techniques for combining information. All of these methods require an external criterion. That is to say, the information about the individual is divided into two distinct parts, one being the collection of information about him which is used in order to make statements about the other one single item, namely, the "external" criterion (e.g., reconviction, return to prison, or other measurement or classification). But there is a different family of techniques which do not begin with the idea of using a body of information to make statements or predictions about an event, classification, or measurement external to the information so used. These methods ask only about the differences between persons or information where it is treated as all alike and not separated into two parts (information and criterion sets).

If we have a large collection of information it is possible to ask how it may be divided up. It can be divided up in terms of two different considerations. We may say that we require similar information to be put together or that we require similar persons to be put together. In these methods we use the data to determine its own subdivision according to certain rules. This family of methods is usually known as "taxonomic analysis."

Again, there are several ways in which the mathematical methods may be applied; and, again, they have a similarity. Instead of examining the body of information in the light of the question, which item is most powerful in prediction of the criterion, we examine the data and ask either (a) which person, in the light of the information about himself and all other persons, is the most unlike all others (a sort of leader of the opposition!); or (b) which information, in the light of all the information, separates the set into parts with the greatest efficiency. If we follow (a) once we have identified the person who is "most different" from all others, we may set him aside and search the remaining persons "set" for persons who are more like him than others. We then "transfer," as it were, the identified individuals until such a time as the differences between the groups are maximized and the difference within the groups are minimized. A similar procedure can be used for items of information.

When the "taxonomic analysis" is completed (without reference to the criterion of reconviction or other such measure), we may examine the different classes of persons identified either by the "person" subdivision or the "information" subdivision to see how well the method discriminates an external criterion such as those who become recidivists from those who do not. At this stage in this project, this is the matter of concern.

Perhaps these methods seem sophisticated and reasonable. Certainly, statisticians have advocated methods of these kinds; and the power of the methods has been demonstrated in many fields of operational research, business, and economic analysis. The argument is convincing both in theory and by example that these methods must be superior to the simple allocation of weights without reference to any of the statistical theory of regression or taxonomy. But, convincing as the theory is and sound as the examples are from other fields of application, the fact remains that in the field of criminological prediction these methods do not work too well. We must, of course, define what we mean by "working well," but perhaps this may be deferred for a moment while we look at some cruder methods of adding information together.

There is a time-honored system of adding information used in the marking of examination papers. Each question assessed as correct is given a mark of one point, and the "score" is the simple sum of the number of correct answers. Another system assesses some questions as more difficult than others and gives weights according to the difficulty which is believed to attach to the questions.

One of the earliest, if not the first, prediction table for the use of parole boards was constructed by Burgess. He used the simple weighting system, giving one point for each item of information about an offender which was associated with later success and one point with a negative sign attached for each item which was negatively associated with success. Items which did not show any correlation with later success or failure were omitted. This is called the "Burgess system" of weighting. It is, without doubt, the simplest possible system of adding information. It makes no allowance for the overlap factor.

Another name associated with parole and other criminological prediction table construction is that of the husband and wife team, Sheldon and Eleanor Glueck. Their method of weighting was to consider the percentage differences between the successes and the failures and to give weights accordingly. Again, there was no consideration of the overlap between items of information.

Both the Burgess system of weighting information and the Glueck system may be used for any number of items of information. However, while Burgess used a large number of items, the Gluecks tried to make the best "prediction" possible with as few items as possible.

There is no statistical theory to support either the Glueck or the Burgess system of weighting. It may, however, be expected intuitively that the Burgess system, if used with a large number of items, would be less likely to be in error than that of the Gluecks with a small number of items and with large weightings given to some two or three facts or assessments.

Prediction tables are constructed by the use of experience of the past. Any method of weighting of information, for the purposes of testing methods for its addition, can be constructed only upon the basis of data from one sample. The sample which is used for "construction" of the tables is, not surprisingly, called "the construction sample." Whether the tables work for the construction sample or not is unimportant to those who wish to use them for the future. Of course, it is unlikely that a good table could be constructed without its fitting the construction sample; but the test of the table is not how well it fits the data upon which it is based but rather how it fits other (i.e., future) data. This would seem to be an obvious fact of considerable simplicity.

Testing the tables on other samples is known as "validation," and the samples upon which it is so tested are termed "validation samples." It might seem surprising that very few prediction tables have been tested on validation samples, but such is the case. Where there are exceptions, of course, it has been noted that the power to predict success or failure on the basis of information (added together by any means preferred by the constructors of the tables) is considerably less for the validation sample than it is for the construction sample.

The difference between the power to separate successes from failures (or any other external criterion) in the construction and in the validation sample is termed "shrinkage." Since hindsight is always more correct than foresight, the power of prediction is always some degree smaller in the validation sample. Correlations between the "scores" (calculated for individuals by the addition of items of information about them) and the criterion are used as measures of the power of the tables, and the differences between the correlations measure the shrinkage. Despite the fact that everybody must have expected shrinkage to take place between construction and utilization or validation, very few of those who have built "prediction tables" have actually tested them in any prediction situation. Moreover, it is only within the last year or two that studies have appeared which have tested the degree of shrinkage which occurs with different methods of building the tables. Statisticians have, it seems, been convinced by the theoretical support for the principles of optimal estimation, while others were not concerned with statistical sophistication and did the best they could with rather simple and intuitively satisfying procedures.

From studies in different countries and with very different sets of data derived from various sectors of the correctional processes, there are now sufficient data to make it clear that, when tested against the hard reality of utilization _in_ a prediction situation, the most efficient statistical methods suffer considerable shrinkage. Often the shrinkage is greater for the more "powerful" methods than for the simple methods of addition, such as that employed by Burgess nearly half a century ago. The several studies which have been published, together with our own data in the present project, may be summarized as follows: the more powerful and efficient the statistical procedures for the addition of information into a prediction score, the better the score fits the "construction" sample; however, when a variety of possible methods are used on one set of data and tested on validation samples, the less powerful methods shrink less and may (indeed, usually do) end up in practice better than the sophisticated techniques.

It was in the light of these results that it was decided to prepare for the use of the federal parole board under this project a broadly based experience table: where the addition of items of information was carried out by means of the simple Burgess weighting system.[1]

There is no doubt that the simple methods have been shown to be trustworthy, while the complex methods of weighting have not. There can be no doubt, too, that the statistical theory is correct. These two statements seem to be in direct conflict. It must be remembered, however, that statistical theory is based on assumptions about the basic data upon which the calculations are made. The phenomenon observed, which may be summarized as "inefficient statistics are best," is true only in respect of data which do not satisfy the assumptions which are made in statistical theory. The peculiar finding does not suggest that there are factors concerning <u>offenders</u>

---

[1]Gottfredson, D.M., Wilkins, L.T., and Hoffman, P.B., <u>Summarizing Experience for Parole Decision-Making, Report Number Five</u>, Davis, California: Parole Decision-Making Project, National Council on Crime and Delinquency Research Center, February, 1972 (draft).

(about whom data are collected) which give rise to this odd result, rather the nature of the data as they are collected, recorded, and classified must provide the clue. Statistical efficiency assumes quality data, and this assumption is not satisfied with respect of the information recorded regarding offenders in the penal system.

Any major advance in the development of prediction methods--and all those other techniques which could be marshalled by research workers through modern technology--must await better quality data. Data of the kind necessary can only be obtained if they are honest at source, carefully recorded, and efficiently transmitted throughout the processes concerned. It is interesting to note that a very considerable amount of money is spent on auditing accounts--even down to trivial detail; but statistics, upon which equally important decisions are based, are subject to no audit or monitoring function. The recording of statistical information is often relegated to the lower levels of clerical worker, and those who would suffer serious pangs of conscience to enter a few pennies under the wrong heading in accounts feel quite free to "adjust" statistical data as seems reasonable to them.

Doubtless some persons concerned with the correctional management system will regard this finding as a blinding glimpse of the obvious. Everybody, it may be claimed, who is closely connected with the processing of offenders knows that the recording of information is not treated with any great respect; and that in some establishments the offenders themselves have some responsibility for some of the recording procedures. To arrive at this result, the research workers, as usual, have gone the long way around and have introduced plenty of inconsequential theory! Perhaps the poor quality of the basic data is obvious to some persons, but those persons presumably use the information recorded, or some of it, to make their decisions regarding disposition of offenders, provisioning, or transportation and other questions. It has, it must be assumed, generally been regarded that the quality of the information was "good enough" for its purpose and that any investment of money to increase the quality of data was unjustified. This is now clearly shown not to be the case. As a temporary measure to accommodate poor quality data, we may apply poor quality methods to the utilization of it, because this strategy provides a better result than that which we can obtain by the use of higher grade methods. There is some analogy with extraction of minerals: high quality ore is needed if powerful methods of extraction are to be used;

poor quality ore can be used in rougher methods of extraction. But data are not natural products over which we have no control; data about offenders are generated within the criminal justice system. The criminal justice system is the "consumer" of these data; and the same system is concerned (or should be) with the quality of the product. The products generated out of data are decisions. Decisions cannot be better than the data upon which they are based, no matter what techniques of handling the data may be employed. The conflict of statistical theory with experience in the practical world of decision-making in criminal justice has revealed a fundamental problem of the quality of the raw material, and it has shown beyond all reasonable doubt that the quality of the basic information is not inconsequential.

END