Journal of Criminal Law and Criminology

Volume 26
Issue 3 *September-October*

Article 5

Fall 1935

Parole Prediction as Science

William F. Lanne

PAROLE PREDICTION AS SCIENCE

WILLIAM F. LANNE

The science of prediction of the probability of success on parole of inmates released from penal institutions is a comparatively recent departure in the field of penology. Although more or less organized efforts had been made in isolated instances to develop a technique for judging in advance the likelihood of a particular delinquent to rehabilitate himself upon release, the first serious attempt to determine whether such prejudgments were feasible appears to be that of Prof. Sam B. Warner,[1] who in 1923 published the results of a study of the records on parole of ex-inmates of the Massachusetts Reformatory. Professor Warner examined the records of 300 parolees who had succeeded, 300 parolees who had violated parole and 86 inmates who were not granted parole but were required to serve their maximum terms within the reformatory. Warner's article is an outstanding example of patient, thorough, fair-minded, scholarly examination of facts. Every statement is carefully weighed and thoroughly temperate, but absolutely fearless. The study investigates not only the factors considered by the Board of Parole in reaching its decisions, but also the potential value of each of sixty-four other items available to the Board but not utilized. The conclusions reached are that of the criteria now in use only *Recidivism* and *Offense* are true criteria, and that the only item not now used, which might be of prognostic value, is the *Alienist's Report.* Every one of the remaining sixty-odd pieces of information, however, is valueless as a criterion of "parolability."

A few months after the appearance of Professor Warner's study, Prof. Hornell Hart published a criticism[2] of the nul findings of the investigation. He attacks Warner's work at its most pregnable point: the fact that no tests of significance of the differences found were applied. Warner appears to have considered that mere inspection of his tables would reveal that such differences as appeared between violators and non-violators were so small as to be entirely attributable

[1]Warner, Sam B., "Factors Determining Parole From the Massachusetts Reformatory," Jour. of Crim. Law and Criminol., 14: 172-207, August, 1923.

[2]Hart, Hornell, "Predicting Parole Success," Journal of Criminal Law and Criminology, 14: 405-414, November, 1923.

to the fluctuations of random sampling. This view Hart attacks vigorously and, indeed, opens his article with the assertion that the percentage of parole violation could be reduced by half through proper utilization of Warner's figures.

Perhaps the two greatest contributions of Hart's study are the facts that he is the first definitely to outline a procedure for the construction of prediction tables—a procedure followed almost to the letter by the Burgess system, and that he definitely recommends the weighting of factors by testing their inter-correlations.

To Prof. Ernest W. Burgess belongs the honor of being the first actually to prepare an experience table which could be used for prognostic purposes. In 1928 he published a study[3] of the records on parole of 1,000 inmates from each of the three penal institutions in Illinois: the Illinois State Penitentiary at Joliet, the Southern Illinois Penitentiary at Menard and the Illinois State Reformatory at Pontiac. The information contained in the official records of the institutions was tabulated under some thirty headings in the case of each of the 3,000 parolees studied. These tentative factors were then tested for association with success on parole, and such of them as showed an appreciable degree of association were retained. Twenty-one factors were finally employed and prediction tables prepared. Predictions as to likelihood of success on parole, based on this study, are presented to the Parole Board for its consideration in the case of every inmate heard by that body.

In 1930 appeared the first of three detailed and exceedingly able investigations in the field of delinquency by Dr. and Mrs. Sheldon Glueck.[4] These studies differ from that of Professor Burgess in two important respects: (1) They do not accept institutional records at face value, but are based on exhaustive field investigation designed to verify every detail of the information used and (2) They employ a small number (from five to seven) of factors, each of which has been demonstrated to have a relatively high degree of association with success on parole rather than a larger number of factors, whose degree of association with success varies. Further, these factors are weighted in accordance with the actual violation rates found to occur in each sub-category, whereas the Burgess factors are not weighted. The studies of the Gluecks represent the most elaborate and painstaking investigations into this subject which have appeared to date.

[3]Bruce, Harno, Burgess, Landesco, "Parole and the Indeterminate Sentence," 1928.

[4]Glueck, Sheldon and Eleanor T., "500 Criminal Careers," 1930; "Five Hundred Delinquent Women," 1934; "One Thousand Juvenile Delinquents," 1934.

The year 1931 saw the publication of two important contributions in the field of prediction: a follow-up study of the Burgess system by Clark Tibbitts[5] and an investigation into parole prediction in Minnesota by Prof. George B. Vold.[6] Tibbitts' study consists of an extension, with certain refinements, of the original Burgess technique to a group of 3,000 consecutive parolees from the Illinois State Reformatory. The important suggestion is contained in this article that categories whose violation rate differs from the mean rate of the group by less than 5% be discarded as insignificant.

Professor Vold's investigation concerns a group of 542 parolees from the Minnesota State Prison and a group of 650 parolees from the Minnesota Reformatory. In addition to the new data presented the study is most valuable in that it is the first published attempt to settle experimentally and by actual appeal to facts such important questions as optimum number of factors, importance of the degree of association with outcome exhibited by the factors employed and the relative merits of weighted and unweighted methods of scoring.

In addition to the published studies quoted, there have been available to the writer an unpublished experience table based upon 2,772 consecutive parolees from the Southern Illinois Penitentiary, covering the men paroled after August 16, 1929, and prepared by Mr. Sam Daykin, and an experience table prepared by Mr. Ferris F. Laune, and based on 1,569 consecutive parolees from the Illinois State Penitentiary, which includes all inmates released on parole from that institution from January 1, 1927, to December 31, 1930. Finally, Mr. Laune has kindly permitted the writer access to an as yet unpublished study which marks a definitely new departure in the field of prediction and bids fair to revolutionize prognostic practices.

It is the belief of the present writer that the data described in the studies enumerated yield an ample basis for an attempt to distinguish between those elements which are of universal application in prediction and those elements which are purely fortuitous or, at most, confined to the specific study in which they appear. In a sense parole prediction has been a fine art; the information seems now to be available to place it upon a sound scientific footing. It is in the hope that some suggestions tending in this direction may be offered that this article is written.

---

[5]Tibbitts, Clark, "Success or Failure on Parole Can Be Predicted: A Study of 3,000 Youths Paroled From the Illinois State Reformatory," Journal of Criminal Law and Criminology, *22*: 11-50, May, 1931.

[6]Vold, George B., "Prediction Methods and Parole," 1931.

Exhaustiveness of Factors

In addition to the question whether the factors actually employed in any study are associated with success on parole and therefore important as elements in a prognostic device, there must arise the further question whether there may not be yet other factors of equal or even of greater importance in this connection. In the cases of Warner, Burgess, Tibbitts, Vold and Daykin this matter has been settled practically by the question of the availability of data. Since these investigators were limited to the information contained in the institutional records, it is obvious that their choice of factors was narrowly circumscribed. A typical expression of this circumstance occurs in Tibbitts' study.[7] He remarks, "In the original study (Burgess' 1927) the committee was concerned less with making an empirical selection of the factors which might be correlated with outcome than with choosing those upon which the records would yield information and which might then be selected through statistical examination." Of the published studies only those of the Gluecks, in which field investigation was used to supplement documentary information, would appear to have contained in their approach any possibility of exhaustiveness of factors. Even they, however, frankly admit[8] that although they have studied some fifty factors, it is not impossible that there may be still others of equal importance.

It is only in the case of Laune's study,[9] with its entirely new approach to this question, that there appears any possibility of determining objectively when the series of significant factors has been exhausted. Laune sets out to investigate the truth of the opinion, very widely held by penologists and prison authorities, that if it were possible to obtain the honest, unprejudiced judgments of an intelligent inmate as to the probability of success on parole of his fellow-inmates, such judgments or "hunches," as Laune prefers to call them, would prove to show a much higher degree of association with actual outcome than any prognostic device yet invented. His first step is to establish in a most satisfactory manner the actual existence and validity of such "hunches" by an analysis of the correlations existing between the "hunches" of various inmate investigators. The correlations found give presumptive evidence in favor of the existence of a definite, measurable quality, which he denominates "parolability." The next step in the research is the isolation of the unit factors which

[7] *Op. cit.*, page 14.
[8] "500 Criminal Careers," page 296, footnote.
[9] Laune, Ferris F., "A Technique for Developing Criteria of Parolability" (to be presented as a doctoral dissertation, Northwestern University).

combine to make up such total parolability. This part of the Laune study is one of the most ingenious means yet devised for objectifying intrinsically subjective material. It consists in the scoring by two inmate investigators of each of a group of 150 subjects by the "hunch" method. Such scoring is done absolutely independently. Next, the scores given by each "huncher" for every individual subject are compared by the "hunchers" themselves, and wherever the slightest difference of judgment appears the investigators indulge in protracted debate, each attempting to defend his own score. A careful stenographic transcript is kept of this discussion and, of course, this permits the actual tabulation of such facts as how early a given factor emerged in discussion, how often it was mentioned, etc. When it is remembered that such discussions took place daily for from two to three hours and over a period of six weeks, it must be admitted, at least, that the method does not lack thoroughness. Laune's warrant for believing that his 54 factors are exhaustive, in the sense that they include all elements which enter into inmates' "hunches," lies in the fact that during discussion of the first 25 subjects new factors emerged so rapidly as to strain the abilities of the stenographer; from the 26th to the 39th subject the rate of emergence drops very steeply; and after the 52nd subject no new factor is found in spite of the fact that the discussion continued with equal thoroughness through the remaining 98 subjects. In this contribution to the subject of exhaustiveness lies one of the most important contributions of this unique study.

Number of Factors

Perhaps there exists in the field of prediction no more moot question than what constitutes the optimum number of factors to be employed in prognostic tables. Burgess and those investigators who have continued his general method of approach hold that as large a number of factors should be employed as are available and can be shown to possess the necessary characteristics: reliability, association with outcome on parole and freedom from linkage or inter-correlation. To these requisites the present writer would add one important item—stability, which will be described below.

The justification for this procedure lies, of course, in the desire to use to the fullest extent all the available information. This point of view is presented very forcefully by Van Vechten,[10] who points

[10]Van Vechten, Cortlandt, C. Jr., "Parole Prediction: Problems of Reliability and Follow-Up" (MS. of a paper delivered before the American Sociological Society, Dec. 29, 1934), p. 2.

out that in addition to the obvious desirability of using the maximum possible information there is the danger that when only a few factors are used, "a single error in classification under a single category assumes relatively great importance, whereas with more factors used the significance of a single error is minimized."

The opposite side of this question: the use of relatively few factors, which exhibit a high degree of association with outcome, has been championed chiefly by the Gluecks. In 500 CRIMINAL CAREERS it seems to be assumed as self-evident that the careful use of a few significant factors constitutes the proper procedure; in 500 DELINQUENT WOMEN this practice is defended on the grounds that (a) the experimental tests of Vold[11] and Monachesi[12] have demonstrated that the use of the Burgess and of the Glueck methods of scoring on their material gives very nearly identical results; (b) the use of but a few factors results in the saving of much time and effort, and (c) being more easily applied, the abbreviated method is likely to prove somewhat more reliable in the hands of inexperienced investigators. In ONE THOUSAND JUVENILE DELINQUENTS the Gluecks do concede the probability of a "slight" gain in accuracy when a larger number of factors is employed, but are ready to sacrifice it on the grounds stated.

Both Vold[11] and Monachesi[12] have put the rival scoring methods to an experimental test by applying both to some of their data and comparing the degree of accuracy of the results. Both report a very high degree of correlation between the two systems and neither ventures an opinion as to the superiority of one or the other method except on the grounds of expediency. This matter will be discussed more fully in a later section of the present paper. Here it is sufficient to state that Vold concludes[13] that more accurate results were obtained when predictions were based on 25 "selected" factors than when prognoses were built up from use of the 17 factors exhibiting the highest degree of association with outcome. This conclusion seems definitely to be in error and to result from Vold's failure to apply any very crucial test of accuracy to his tables. A criterion for use in such cases—the mean probability, per individual, that the observed differences between predicted and actual rate of violation are due to chance—is discussed briefly in a later section.

In reply to the contention of the Gluecks it seems only fair to say that in a matter where the protection of society hangs in the balance on the one hand and a portion of the life of a large number of

[11] *Op. cit.*, page 95.
[12] Monachesi, Elio G., "Prediction Factors in Probation," 1932, page 108.
[13] *Op. cit.*, page 100 and page 105.

individual delinquents on the other, any sacrifice of the convenience of the investigators would appear to be justified if even a slight increase in accuracy of the resulting prediction table can be demonstrated to result therefrom.

On the other hand, the indefinite multiplication of factors leads, unless safeguards are thrown about the procedure, to what may be termed the "single group fallacy." Let us consider any single finite group of paroled inmates upon each of whom there exists complete information and whose individual outcome on parole is also known—the situation, approximately, which exists in all the studies considered. Any unit character whatsoever, whether etiologically related to outcome or not, can be tested for association with success on parole. Some characters will exhibit a relatively high degree of correlation, positive or negative; others will manifest only a very low degree of correlation. In fact, as the number of factors tested approaches infinity the curve representing the degree of correlation with success of the unit characters will approach smoothness. In the case of those factors which are, in fact, uncorrelated with parole success, the coefficient of correlation would tend, as successive groups were tested, to fluctuate about the mean value of zero. In any single group, however, it would be extremely unlikely that the value obtained for the correlation coefficient would be exactly zero. Such deviations from the mean value of zero are, of course, merely errors of random sampling. Yet, in any *one particular group,* it is possible to consider a greater and greater number of factors until, theoretically at least, we should have but one of our subjects in each category: *i. e.,* our factor score would be a perfectly definite "photograph" of some one individual in the group studied. If, for instance, to the factors of *Weight* and *Height,* utilized by Vold, we were to add *Color of Eyes, Color of Hair, Shape of Finger-Nails,* etc., we should eventually reach the point where each individual's factor score would constitute a perfect means of identification of just that particular individual. In any one finite group, then, it is a foregone conclusion, since the outcome of each individual is known in advance, that as the number of factors employed approaches infinity the correlation between prediction and outcome must approach unity. But the whole process becomes absurd at a certain point. For, if what was originally desired was merely to know the actual outcome of each individual *qua* individual, it would have been much easier merely to read this information from the records.

And, of course, any such scheme would not constitute a predic-

tion table for precisely the reason that it would fail to predict on any group other than the particular one studied. The entire essence of prediction theory rests upon the hypothesis that it is possible to isolate certain factors which are universally, or at least very generally, associated with outcome on parole. It is this search for universals underlying the particular phenomena in any experiment or observation which differentiates science from mere description.

In an attempt to establish parole prediction methodology as very definitely a branch of science the author ventures to suggest the following four requisites which, in his opinion, must demonstrably form part of any valid prediction factor:

(1) Reliability
(2) Significance of association with outcome
(3) Freedom from inter-correlation with other factors
(4) Stability

Each of these pre-requisites will now be defined individually. It may be stated at this point, however, that the author believes that there exists in parole prediction a perfectly definite optimum number of factors, and that this optimum number must be defined as the maximum number of factors which possess the four pre-requisites mentioned.

Reliability

It is a truism that a prognostic instrument cannot possess a higher degree of reliability than is manifested by the data from which it is composed and the methods by which it is compiled. And yet there appears to exist a widespread tendency to utilize data in the face of certain knowledge of their unreliability. Fortunately there exist several studies designed to measure the degree of reliability quantitatively. As regards the reliability of the data themselves the most important study is certainly that of Sutherland and Van Vechten.[14] In this paper the degree of unreliability is stated in terms of the percentage of inconsistency. This percentage is found to vary from 0.8% for *Race* to 69.7% for *Occupation.* Warner examined fifty complete files to determine how reliable were the answers of inmates to a questionnaire administered upon admittance to the institution. He finds discrepancies in a quarter of the cases.[15] Laune states that the methods employed in eliciting information from in-

[14]Sutherland, E. H. and Van Vechten, C. C., Jr., "The Reliability of Criminal Statistics," Journal of Criminal Law and Criminology," 25: 10-20, May-June, 1934.
[15]*Op. cit.*, pages 193-194.

mates at his institution are highly unreliable. In this respect the work of the Gluecks, with its verification of all official information, far transcends any other in the field.

But even waiving the reliability of the original data, there still remains the matter of the reliability of the classificatory methods employed. In this field there is the important paper of Tibbitts[16] in which is measured the reliability of the classifications in the cases of 907 inmates paroled from Pontiac. Tibbitts had classified these cases in 1927 for the purpose of the Burgess study; in 1928 he independently reclassified the same cases. He finds the coefficient of reliability for the total factor score to be + .763 ± .009. He analyzes this result and finds that the unreliability is accounted for almost entirely by the difficulty of classification rather than by clerical errors. He continues to an analysis of the reliability of the individual factors and finds eight: *Sentence, Offense, Age, Nationality, Lesser Plea, Time Served, Previous Criminal Record* and *Punishment Record* satisfactory from the point of view of this criterion.[17] The factors found unreliable are: *State's Attorney's Statement, Personality Type, Type of Offender, Size of Community, Mobility, Associates, Mental Rating, Work Record* and, by implication, *Social Type.*[18]

Van Vechten concurs[19] in the belief that the errors of classification are due almost entirely to real variations in judgment and not to clerical errors. He finds that only about 10% of the observed variance can be attributed to the latter. Vold devotes an entire chapter[20] to the careful investigation of the matter of reliability. His reliability coefficients vary from + .205 in the case of *Social Type* to + .925 in the case of *Parole Community* and have a mean value of + .713. All investigators agree that reliability varies directly as the objectivity of the concept underlying the various factors.

It must be obvious that the inclusion in a prediction table of factors whose reliability is but .20 or .30 cannot but play havoc with the final results. Conceding that *Social Type,* for example, may actually be very highly associated with outcome on parole, yet, if we are unable to develop a consistent method of measuring the former, its use as a component element in a prognostic device must be almost nil. It is the suggestion of this writer that factors intended for predictive use be rigidly limited to those in which a comparatively high degree

[16]Tibbitts, Clark, "Reliability of Factors Used in Predicting Success or Failure on Parole," Jour. of Crim. Law and Criminol., 22: 844-853, Mar., 1932.
[17]*Ibid.,* page 848.
[18]*Ibid.,* pages 851-852.
[19]*Op. cit.,* MS, page 5.
[20]*Op. cit.,* pages 58-82.

of reliability, such, for instance, as would be expressed by a coefficient of reliability of .60 or at least .50, can be shown to exist.

Significance

The significant association of a given factor with outcome on parole would appear to be so obviously a *sine qua non* of its use as to render discussion unnecessary. And yet, this seems not to be the case, for in some cases factors have been used, particularly by followers of the Burgess technique, some of the categories of which, at least, are so evidently not associated with outcome that it is difficult to understand why they were even considered. The explanation probably lies in the fact that the total scores of individuals are usually expressed in terms of number of favorable factors. With such a scoring device it is, of course, imperative that every subject be scored on every factor; otherwise the scores of different individuals would not be comparable. This unquestionably accounts not only for the retention of sub-categories whose violation rates do not manifest significant deviations from the mean, but also of the ubiquitous category, *No Information.* The obvious absurdity of using such a category in a prediction table is well exemplified in Laune's findings as regards *Wassermann Test.* In his group of 1,569 inmates it was unfavorable to have either a positive or a negative Wassermann reaction; the only favorable circumstance was to have no record of a blood test on file! Cases occur, however, in which the absurdity is not so apparent. There does not appear to be any *a priori* reason why the category *Robbery* under the factor *Offense* should not possess an equal degree of significance with *Sex Offenses.* And yet, as will be shown in the discussion of stability of factors, the latter is a very highly reliable and significant category, being very distinctly favorable in every study yet reported, while the former appears to lie so close to the mean of the whole group that the chances are only about four to one that it will prove a favorable category in any one given group. A device for avoiding the dilemma engendered by the "number of favorable factors" system of scoring will be outlined.

The logical conclusion seems to be that in parole prediction as in other quantitative disciplines factors should be tested for significance by the usual means and such categories as are found to have critical ratios of appreciably less than 3 should be discarded. A qualification of this rigid criterion, based upon the cumulative significance of a series of differences, will be discussed under stability. The labor of computation of the individual critical ratios for all the

sub-categories employed in a given study is sometimes very great. To obviate this Tibbitts suggests[21] a "rule of thumb" which probably represents a sufficiently good first approximation. He suggests that all categories be discarded as insignificant, which do not differ from the mean violation rate by 5%. The only revision of Tibbitts' suggestion which the author desires to make is that 5% be retained as the critical value for groups in which, as in Tibbitts', the mean violation rate is in the neighborhood of 25% (and in cases where it is about 75%), but that in studies like Warner's, where the violation rate is nearly 50%, the critical value be raised to 6%.

The failure to ascertain that categories are really significant is, of course, the basis of the single group fallacy, described above. It is, perhaps, not out of place to reiterate that in any one particular group the coefficient of correlation between factor score and outcome can be made to approach unity through an indefinite increase in the number of factors used, but that this correlation is spurious and has no general significance whatsoever.

Orthogonality

That good prediction factors must be free from inter-correlation is a maxim which has been repeated in nearly every study since Hart's, but observed, apparently, only in the breach! For example, Burgess employed[22] as two separate factors *Type of Criminal,* with the sub-categories (a) First offender, (b) Occasional offender, (c) Habitual offender, (d) Professional criminal; and *Previous Criminal Record,* with sub-categories (a) No criminal record, (b) Industrial school record, only, (c) Fine or probation, only, (d) Workhouse or jail record, only, (e) Reformatory record, (f) Penitentiary record. Since a "first offender" is, presumably, an individual with "no criminal record" and an "habitual offender" and a "professional criminal" would very probably have reformatory or penitentiary records, it is difficult to believe that the two factors mentioned could have other than a high degree of correlation. With minor changes, which seem rather to aggravate than to mitigate this defect, these same two factors are retained by Tibbitts, Daykin and Laune. This would seem to amount to the very definite weighting of the common factor involved. Another pair of factors which, on their face, seem to be highly correlated are *Offense* and *Sentence* (employed by Burgess, Tibbitts, Vold, Daykin, Laune). Even the Gluecks, with their rela-

[21] Tibbitts, Clark, "Success or Failure on Parole Can Be Predicted," etc., Jour. of Crim. Law and Criminol., 22:44, May, 1931.

[22] *Op. cit.,* pages 224 and 228.

tively few highly-associated factors, fall into this pitfall, for of the six pre-reformatory factors used for prediction in 500 CRIMINAL CAREERS, it is difficult to believe that (1) *Seriousness and frequency of crime,* (2) *Arrest for crime preceding the offense for which sentence to the reformatory was imposed,* and (3) *Penal experience preceding reformatory* are all mutually orthogonal, or that (4) *Industrial habits* and (5) *Economic responsibility* are strictly independent. In ONE THOUSAND JUVENILE DELINQUENTS, it is at least possible that *Discipline by father* and *Discipline by mother* are inter-correlated.

An interesting speculation in this regard is that inter-correlation is very probably the explanation of the association with outcome exhibited by some factors for which it is difficult to postulate direct etiological relationship. For example, Vold finds both *Height* and *Weight* to be positively correlated with success on parole, the degree of association being expressible by coefficients of mean square contingency of .086 and .069 respectively.[22a] When it is remembered that Vold's highest *C* is but .283, this amounts to an appreciable degree of association. Yet it is difficult to see at first blush why tall and heavy men should be more likely to succeed on parole than short and light men. An explanation, which is, at least, possible, lies in the facts that (a) it is significantly favorable in Vold's group to have been brought up on a farm,[23] (b) of Minnesota's large Swedish population a considerable portion is rural, (c) Swedes are, presumably, on the average both taller and heavier than the mean of the population as a whole, therefore, *ceteris paribus,* (d) it is favorable in Vold's group to be tall and heavy. This explanation, while perhaps far-fetched, is at least somewhat less so than the familiar connection between the clover crop in a given area and the number of old maids resident in that area.

The entire subject of inter-correlation of factors is badly in need of investigation. The writer hopes to study this question in Laune's group of 1,569 subjects.

Stability

A final criterion of what constitutes a good prognostic factor rests upon a concept which the writer does not find mentioned elsewhere in the literature on the subject. This is the concept of stability of favorableness or unfavorableness.

The usual statistical criterion of significance: that the ratio of

[22a] *Op. cit.,* page 56.
[23] *Ibid.,* page 42.

an observed difference to its standard deviation be at least 3 is, perhaps, somewhat too severe in the instance under discussion; rigid adherence to it may cause the elimination of genuinely valid information. Further, as Peters has pointed out,[24] no magic significance is attributable to just this particular critical ratio. It rests, in the last analysis, merely upon convention. More important is Peters' second point:[25] that the statistical significance of a *series* of differences exhibiting a definite trend is quite different from the question of the significance of the differences considered individually.

An analysis of the twelve different groups of delinquents considered in this paper offers a splendid opportunity for the definition and evaluation of a "criterion of stability"—a perfectly definite, quantitative measure of the probability that the favorableness or unfavorableness of a given sub-category is significant and not due to random fluctuation. The procedure employed is as follows: for a given sub-category of any factor the violation rates reported in each of the twelve studies are classified as (a) Significantly favorable, (b) Insignificantly favorable, (c) Insignificantly unfavorable, (d) Significantly unfavorable, using as a criterion the rule of thumb quoted above: that those categories shall be considered significant, whose violation rates differ from the mean violation rate by more than 5% (or by more than 6% in the case of groups where the mean rate approximates 50%).

Now the possible range of the probability that an observed difference is not due to chance is from .50 (for an observed difference of zero) to a value approaching unity as a limit (.999 for a difference of 3σ). Let us take as a rough first approximation the limits of the inter-quartile range and assign, somewhat arbitrarily, to the categories classified as significant the probability .875 that the difference is not due to chance, and to the categories classified as insignificant a corresponding probability of .625. Subtracting these probabilities from unity we have a probability of .125 that the categories labelled significant represent a random fluctuation, while the corresponding probability for the insignificant categories is .375.

We next determine the general trend of the sub-category under consideration: by inspection we estimate whether it is, on the whole, favorable or unfavorable. Let us assume that such inspection shows a given category as, on the whole, favorable. We count the number of groups of inmates in which the category was insignificantly favor-

[24]Peters, Charles C., "Note on a Misconception of Statistical Significance," American Journal of Sociology, 39: 231, September, 1933.

[25]*Ibid.*, page 234.

able and the number in which it was significantly so. Using these two numbers (let us say w and x) as exponents, we find the product of $(.375)^w$ by $(.125)^x$. Next, we determine the number of cases in which the category was found to be insignificantly and then significantly unfavorable. Denoting these numbers by y and z, we set up a fraction of the form

$$\frac{(.375)^w x\ (.125)^x}{(.375)^y x\ (.125)^z}$$

This fraction evidently represents the probability that the whole series of observations yields a general trend which is merely a random fluctuation. I propose that this value be known as the index of instability, and be represented by the letter v.

By way of example we shall now quote several actual computations of v, the first in the case of a very stable sub-category: that of *No previous convictions* under the factor *Previous Criminal Record.* In tabular form this appears as follows:

PREVIOUS CRIMINAL RECORD

No Convictions

Warner	*Burgess*			*Tibbitts*
	Joliet	*Menard*	*Pontiac*	
S.F.	S.F.	S.F.	S.F.	S.F.

Gluecks			*Vold*		*Daykin*	*Laune*
Ref.	*Juv.*	*Women*	*Pen.*	*Ref.*		
S.F.	S.F.	S.F.	S.F.	S.F.	S.F.	S.F

Here the trend is, of course, favorable. Since in each of the twelve groups studied the category is significantly favorable, the probability that the entire series of observations is due to chance is expressible by the value $(.125)^{12}$. In other words, $v = 0.00000000001455$ or $(1.455) \times 10^{-11}$. This factor, then, is remarkably stable; it is a certainty that the series of results obtained is not due to chance.

In the case of the category *Urban* under the factor *County from which Received,* the findings are quite to the contrary. Only four of the studies analyzed contain information on this category, as follows:

COUNTY FROM WHICH RECEIVED

Urban

Burgess	*Vold*		*Laune*
Joliet	*Pen.*	*Ref.*	
I.F.	S.F.	I.U.	S.U.

Here the findings exactly balance each other and $v = \frac{(.375)}{(.375)} \frac{(.125)}{(.125)} = 1$, which means that on the basis of these findings the category is totally uncorrelated with success on parole and that the chances are exactly even in any given study that its violation rate will be greater or less than the mean rate.

This is the case, also, in the category *23-24* under the factor *Age*. The findings are as follows:

AGE

23-24 Years

Warner	*Burgess*			*Tibbitts*
	Joliet	*Menard*	*Pontiac*	
I.F.	S.F.	I.F.	I.U.	I.U.

Gluecks	*Vold*		*Daykin*	*Laune*
Ref.	*Pen.*	*Ref.*		
I.U.	S.U.	I.F.	I.U.	I.F.

Here $v = \frac{(.375)^4 \times (.125)}{(.375)^4 \times (.125)} = 1$, and hence this category, too, seems to be quite meaningless for parole prediction.

In all, 312 separate sub-categories, reported by eight investigators and based on twelve different groups, were studied. The values of v were found to range from 1.5×10^{-11} to unity. Of these values 30.4% represent a probability less than one in a hundred that the results are mere random fluctuations; 23.1% represent a corresponding probability less than one in a thousand. It must be remembered, of course, that the value of v varies inversely as the number of investigators reporting on the category. Since some categories are reported by only a few or even by but one investigator, there is a

probability that some categories found to have high values of v will exhibit, instead, low values when more results are in.

For the present, however, the writer desires to suggest that only those sub-categories be used in prediction tables whose index of instability is less than .01. As more studies appear and particularly as there occurs an increase of uniformity among various investigators in the categories studied, very many other categories, whose index is now in excess of this value, will unquestionably be found to be valid. It is recognized, of course, that greater accuracy could have been obtained in these computations by calculating the ratio of difference to standard deviation of the difference in each of the two thousand separate entries considered and then reading from a table of integrals of the normal probability curve the chances in a hundred that the difference is due to chance, rather than using arbitrarily the values .375 and .125. Lack of time determines our choice here. We hope to perform the computations later.

It is the very earnest suggestion of the author that all investigators in the field of parole prediction cooperate in the building up, as rapidly as possible, of a corpus of data which will be highly reliable. This can, perhaps, best be done by the prompt reporting on the part of every investigator of his findings with respect to every new category employed by him. All other investigators who have the opportunity of doing so should then use identical categories in their groups and report their results. In this manner it will appear very soon whether the new categories or factors are significantly correlated with outcome or not. As the initial move in this program the author appends the following list of the factors studied by him, together with their respective sub-categories and the values of v for each.

Factor	*Trend*	v
1. AGE		
15	U	.0469
16	U	.1250
17		1.0000
18		1.0000
19	U	.0176
20	U	.0025
21	U	.1250
*22	U	.0176
*22 or less	F	.0469
23-24		1.0000
25-26	U	.0001
27-28	U	.0022
29-30	U	.0025
31-32	U	.3750
33-34	U	.0527
35-36	F	.3333
37-40	U	.1582
41-45	F	.0469
46-49	F	.0059
50 and over	F	.0007
2. ASSOCIATES		
None	U	4.88×10^{-5}
One	F	.0176
Two	F	.0025
Three or More	F	1.14×10^{-5}
Not Here and Not Apprehended	U	.1250
3. CHURCH ATTENDANCE		
Regular	U	.0020
Irregular	U	.3333

Factor	Trend	v
4. Community When Arrested		
Metropolitan	U	.0176
Urban	U	.0020
Towns	F	0007
Villages	F	.0008
Rural	F	3.81×10^{-6}
No record	U	.0020
5. County From Which Received		
Metropolitan	U	.0156
Suburban Metropolitan	U	.3333
Urban		1.0000
Rural	F	.0156
6. Family		
Average	F	3.43×10^{-5}
Broken	U	.0008
Left Home	F	.3750
Inferior	U	.8889
Deteriorated	U	.0156
Superior	F	.1250
Institution	U	.0020
Parents Living Together		1.0000
Parents Living Separate		1.0000
Parents Living Divorced	U	.0469
Father Dead		1.0000
Mother Dead	U	.1406
Both Parents Dead		1.0000
No Record	F	.0003
7. Intelligence		
A	F	.0022
B	F	.1250
C		1.0000
D	U	.0059
E	U	.0059
No Record	F	.0059
8. Job on Parole		
Farm Regular	F	.1250
Farm Adequate	F	.1250
Regular	F	.1250
Adequate	F	.1250
Physically Unable to Work	F	.1250
Farm Inadequate	U	.1250
Inadequate	U	.1250
Unemployed	U	.1250
Skilled	F	.1250
Semi-skilled	F	.3750
Unskilled	U	.3750
Farm	F	.3750
Business for Self	F	.1250
None	U	.1250
No Record	U	.1250

Factor	Trend	v
9. Lesser Plea		
Yes	F	.0469
No	U	.1406
10. Marital State		
Single	U	.0003
Married	F	.0003
Separated	U	.4219
Divorced	U	.1406
Widowed	F	.1250
11. Mobility		
Resident	F	.0001
Transient	U	5.36×10^{-7}
No Record	U	.3750
12. Months on Parole		
1- 6 months	U	.0469
7-12 months	U	.0469
13-18 months	U	.3333
19-24 months	F	.1406
25-30 months	U	.3333
31-36 months	F	.0469
37-42 months	F	.0469
43-48 months	F	.0469
49 and over	F	.1250
None	U	.1250
13. Nationality		
American White	U	.3750
American Colored	U	3.86×10^{-5}
British and Canadian	U	.0370
German	F	.0066
Irish	U	.0006
Other Nordic	F	1.44×10^{-5}
Slavic	F	.0066
All Others	F	4.29×10^{-6}
14. Neighborhood		
Residential	F	4.29×10^{-6}
Industrial and Immigrant	F	.0176
Rooming House	U	1.14×10^{-5}
Apartments	U	.3333
Farms	F	.0156
Underworld	U	3.05×10^{-5}
Hobohemia	U	.0007
Negro	U	.0156
No Record	U	.0417
15. Number of Violations		
None	F	.3750
One or More	U	.1250
16. Offense		
Homicide	F	.3333
Robbery	F	.2963
Burglary	U	1.61×10^{-6}
Larceny	U	3.86×10^{-5}
Fraud	U	7.94×10^{-8}
Sex Crimes	F	9.31×10^{-10}
Other	F	.0001

Factor	*Trend*	*v*
17. PAROLE COMMUNITY		
Out of State	U	.0417
Metropolitan	U	.0417
Urban	U	.3750
Small Towns and Farms	F	.0007
No Record	U	.0176
18. PAROLE NEIGHBORHOOD		
Rooming House	U	.0156
Apartment House	F	.3750
Immigrant	U	.3750
Negro	U	.0469
Residential	F	.0156
Farm	F	.0156
Small City	U	.3750
19. PERSONALITY TYPE		
Egocentric	U	.0022
Inadequate	F	.0198
Unstable	U	.3750
Mental Defective	U	.3750
Emotional	F	.0001
No Gross Fault	F	.0002
Sexual Psychopath	F	.1250
All Others	F	.0417
20. PHYSICAL DEFECTS		
None	F	.0527
Major Defects	U	.0020
Minor Defects	U	.0156
21. PREVIOUS CRIMINAL RECORD		
No Convictions	F	1.46×10^{-11}
One Conviction	U	5.96×10^{-8}
Two Convictions	U	9.31×10^{-10}
Three Convictions	U	3.49×10^{-10}
Four or More Convictions	U	1.79×10^{-7}
Police Character Only	F	.0022
Fine or Probation Only	F	3.86×10^{-5}
Industrial School Only	U	4.77×10^{-7}
Jail Only	U	1.99×10^{-8}
22. PUNISHMENTS		
None	F	8.38×10^{-9}
One	U	.0002
Two	U	4.77×10^{-7}
Three	U	1.43×10^{-6}
Four-Five	U	1.43×10^{-6}
Six or More	U	4.77×10^{-7}
Demerits	U	.0527
1-2 Demotions	U	3.81×10^{-6}
More Than 2 Demotions	U	.0002
No Record	U	.0469
23. RECOMMENDATION OF STATE'S ATTORNEY		
Recommends	F	1.14×10^{-5}
Protests	F	.3333
Factual	F	.3333
No Record	U	.0001
24. RELIGION		
Catholic	U	.1406
Protestant	F	.3750
Hebrew	F	.0156
Other	F	.0156
25. SENTENCE		
Indeterminate to 5 Years	U	.0469
1 to 10 Years	U	.1406
1 to 14 Years	F	.1406
1 to 20 Years	U	1.29×10^{-5}
1 to Life	F	.0001
10 to Life	F	.0001
3 to 15, 3 to 20	F	.0022
5 to 20 Years	U	.0156
Determinate to 14 Years	F	3.81×10^{-6}
14 to 25 Years	F	3.81×10^{-6}
26 Years to Life	F	5.96×10^{-8}
26. SOCIAL TYPE		
Gang	U	.1250
Mean Citizen	F	.0007
Ne'er-do-well	U	1.14×10^{-5}
Social Inadequate	U	.1406
Sex Pervert		1.0000
Drug Addict	U	.0156
Drunkard	U	.0007
Farm Boy	F	3.05×10^{-5}
Marginal	F	.0469
Hobo	U	9.15×10^{-5}
Immigrant	F	.0022
Black Sheep	F	.3750
Criminal by Accident	F	.0156
Respected Workman	F	.0156
Irresponsible Young Blood	F	.0469
Small Town Bully	U	.1406
Habitual Criminal	U	.1250
Average Negro	F	.1250
Inferior Negro	U	.1250
All Others	U	.0176
27. TIME TO DO ON MAXIMUM		
1 Year or Less	F	.1250
2 Years	U	.3750
3 Years	F	.3750
4 Years	F	.3750
5 Years	U	.3750
6 Years	U	.3750
7 Years	U	.3750
8 Years	U	.3750
9 Years or More	F	.3750
Life	F	.3750

Factor	Trend	v
28. TYPE OF OFFENDER		
First Offender	F	5.96×10^{-8}
Formal First Offender		1.0000
Occasional and Habitual	U	4.29×10^{-6}
Professional	U	.0020
29. USE OF ALCOHOL		
Abstinent	F	.0469
Moderate	U	.0527
Intemperate	U	.0527
30. WASSERMANN TEST		
Positive	U	.0156
Negative		1.0000
No Record		1.0000
31. WORK RECORD		
Regular	F	1.46×10^{-11}
Irregular	F	.0052
Casual	U	6.79×10^{-7}
None	U	1.04×10^{-5}
No Record	U	.3750
32. WORKING WHEN ARRESTED		
Yes	F	.0025
No	U	.0003
No Record	U	.0059
33. YEARS SERVED		
1 Year	F	5.96×10^{-8}
2 Years	F	.0001
3 Years	U	.0003
4 Years	U	2.12×10^{-7}
5 Years	U	4.23×10^{-7}
6 Years	U	1.59×10^{-7}
7 Years	U	8.14×10^{-5}
8 Years or More	U	.2963

*Under AGE, the categories from 15 to 22 refer to juvenile institutions; those beginning with 22 or less, to adult institutions.

Scoring Methods

The relative merits of weighted and unweighted systems of scoring have constituted a controversial point for many years. The Burgess system consists in giving a unit score for each factor in which the individual is classified under a sub-category whose violation rate is less than the mean rate and entirely without regard to the magnitude of the deviation from the mean. The Glueck system, on the other hand, consists in adding in the actual violation rate to three decimal places of every category under which an individual is classified and using the sum as a "failure score" for the individual.

Vold attempted to test the accuracy of the two systems by applying both to some of his data. First he calculated coefficients of correlation between the two systems on his prison group, his reformatory group and the combined group. He found *r's* of + .922, + .925 and + .922 respectively.[26] This, of course, demonstrates the very close similarity of the results of the two systems, but it does nothing to show which is the more nearly accurate. This question Vold investigates by scoring the "operating group" in each of his three units by both methods and comparing the results with the "predictions" based on the "control group." Unfortunately, Vold nowhere defines a satisfactory criterion of accuracy of prediction, basing his opinions as to which is the best of several tables on the consideration that it distinguishes most clearly between the extremes of the range. The author ventures to suggest as a more accurate criterion the "mean

[26] *Op. cit.*, page 95.

probability, per individual, that the observed differences are due to chance," a constant which he has described elsewhere[27] and denominated ψ. When the ψ is applied to the six relevant tables from Vold, the following results are obtained: comparison of the prison "operating group," based on average violation rate for Prison and Reformatory, and scored by the Glueck system (Table XLVIII) with a similar table (Table XL), based on the "control group," yields a ψ of .789. A comparison between the latter and the corresponding table scored by the Burgess system (Table XXXVIII) gives a ψ of .663. In this case, then, the results given by the Burgess method are more nearly accurate by the ratio, roughly, of 4 to 3. In the case of the reformatory group (Tables XLIX and XLI) the comparison is, once more, favorable to the Burgess system, but only slightly so. The ψ's are .739 for the Burgess system and .750 for the Glueck. In the case of the combined group, however (Tables XXXIII and XXIX) the advantage lies with the Glueck method, again, however, only slightly. This time the ψ's are .665 for the Glueck method and .696 for the Burgess. These are the only data given by Vold in such a manner as to make application of the ψ procedure to the two scoring methods possible, and from this limited test it is almost impossible to draw a conclusion, as the advantage is so equally divided.

Monachesi, too, reports comparison of the two systems.[28] He finds a coefficient of correlation of + .865 between the Burgess and Glueck methods on his material. He goes farther, however, and experiments with a method of weighting the Burgess-type factors. When this scoring method is compared with the unweighted Burgess form, the correlation is + .885.[29] He concludes, on reasoning identical with Vold's, that the unweighted method of scoring gives the better results. Unfortunately, Monachesi's data are not presented in such form as to make possible application of the ψ technique.

[27]Lanne, William F., "A Method for Determining the Relative Accuracy of a Series of Prediction Tables." (Under consideration for publication by the Journal of the American Statistical Association.)

The steps in the computation of ψ are as follows: the standard deviation of the difference between predicted rate of violation and observed rate is computed for each category in the table. Next, the successive ratios of observed difference to standard deviation of the difference are calculated. From a table of integrals of the normal probability curve are read off the probabilities that each successive difference is due to chance; these probabilities are multiplied by the number of individuals comprised in the category to which each probability refers; the products are summed, and the sum is divided by the number of individuals in the entire group. This value, then, represents the mean probability, per individual, that the observed differences in violation rates (as between prediction and outcome) are due to chance. That one of several prediction tables represents the most nearly accurate prediction, for which the value of ψ is a minimum.

[28]*Op. cit.*, page 108.

[29]*Ibid.*, page 106.

The Gluecks defend the retention of their method of using a few weighted factors on the basis of the high correlations just quoted.[30] In addition to the objections already adduced it must be added that this argument is not entirely convincing because the comparisons of Vold and Monachesi were not based upon a few weighted factors on one hand and a large number of unweighted factors on the other, but upon an equal number of factors, first weighted, and then unweighted. The high *r's,* then, indicate that the results are altered but slightly when a constant number of factors are weighted. This is a long cry from demonstrating that a smaller number of factors would give equally valid results.

It would appear logical, in advance of further experiment, that the optimum system would be that in which the largest number of demonstrably significant factors were suitably weighted, thus combining the advantages of the two systems.

Either of the systems described, however, renders imperative the scoring of every individual on every factor. So long as this is the case it is necessary to retain categories which are clearly not significant and also the *No Information* category, which cannot be defended on a conceptual basis. As a means of avoiding this necessity the following scoring method is proposed: Let the maximum number of factors and of sub-categories under each factor be retained which meet the four necessary conditions: reliability, significance, orthogonality, stability. Let these categories be weighted, as in the Glueck system, by adding in the percentage of successes[31] found from experience tables to be characteristic of each. Let these scores, however, be summed in two separate groups: the first to consist of the success scores of all those factors on which the individual's success scoring is higher than the mean; the second of the failure scores of those on which the score is lower than the mean. Finally, let the latter be subtracted from the former to give the final score. This amounts, in the unweighted systems, to subtracting the number of unfavorable factors from the number of favorable factors to give a final score. Some of the scores arrived at by this procedure will, obviously, be negative. If this is objectionable, it is an easy matter merely to add a constant to each score, thus shifting the entire scale and making all scores positive.

Causality as a Statistical Concept

There remains one point, not specifically pertinent to the matter

[30] "Five Hundred Delinquent Women," page 287.

[31] "Success score" is chosen in place of "failure score" employed by the Gluecks in the belief that it rests on a more natural conceptual basis.

of parole prediction, but affecting the entire matter of the statistical approach in social science or, indeed, in science in general. It is repeated constantly on all sides that a statistical study cannot hope to be more than a description of observed phenomena—entirely useless as a means of discovering or testing etiological relationships. This point of view the author vigorously challenges.

Dr. Alexander's recent paper[32] constitutes a typical example of the attitude with which the author desires to disagree vehemently. The paper seems not to deal with any but the most elementary statistical techniques. Partial correlation, for example, is not even referred to as a means of holding constant one or more variables while examining the relations of others. But it is not the author's purpose to attack Dr. Alexander's reasoning in detail; rather is it his desire to insist that the statistical approach is not only a valid technique, but that it actually constitutes the only valid technique in science. This perhaps paradoxical statement is based upon two considerations: the psychological basis of the entire concept of causality, and the experimental findings of modern physics, which appear to indicate that the regularity of macroscopic physical "laws" is nothing more than the extraordinary stability of certain statistical means, and that when we actually approach the ultimate physical units—electrons, for example, the only "law" which remains valid is the law of probability.

Considering the widespread circulation of the writings of David Hume and the great following of the English Empirical School, it should not be necessary to remind anyone that he has never actually seen a cause cause an effect. If he believes, with the great majority of scientists, that all knowledge is derived initially through perception, the investigator will realize that he has never once perceived the relationship of causation. And yet he seems to have a definite concept of causality. What has actually been observed is, of course, an invariable relationship of succession between event A (which we may label "cause" if we like) and event B (which we shall call "effect"). But invariable succession is not causation; it is, in fact, a purely statistical relationship. What Dr. Alexander means by "cause" is merely that in the experience of the individual or of the race a very large number of observations have been made in which event B followed event A, while no contrary cases have been observed. This does not, of course, give warrant for the assumption of certainty that such succession will in every case continue to be observable. It does, however, and on very sound statistical grounds, constitute a

[32]Alexander, Franz, "Evaluation of Statistical and Analytical Methods in Psychiatry and Psychology," Am. Jour. of Orthopsychiatry,4 : 433-448, Oct., 1934.

probability approaching unity (certitude) as a limit. As Hume long ago pointed out, the concept of cause is based on an "inferential leap."

The advances made in physical knowledge during the past thirty years constitute an even more unanswerable argument for the conception of causality as statistical in nature. In nineteenth century physics, while attention was concentrated upon phenomena which involved trillions and quadrillions of individual electrons, it was customary to speak of rigid, invariable, causal laws. The first difficulty arose in the case of the Second Law of Thermodynamics, which states that physical processes move in the direction of increasing entropy. But how shall we define entropy? It is apparently a measure of the "random element" in the constitution of the universe. It would be difficult to find a concept more completely statistical! Then came the Kinetic Theory of gases. No physicist today doubts the essential truth of this theory; calculations based upon it enable experimenters to predict results within the limits of observational error and with absolute uniformity. Yet under this theory temperature *is* the mean kinetic energy of the molecules of the substance under observation. Further, calculations are based on the assumption that in the enormous molecule population in a given volume all possible velocities are actually represented in strict accordance with the tenets of the theory of probability. The "invariability of physical laws," then, turns out to be nothing more than the extraordinary stability of the statistical mean of an indefinitely large population!

But in recent years, in experiments actually involving small numbers of electrons, investigator after investigator has announced that definite causation is an empty concept: that all that actually exists in the world of electrons is a probability that an electron will occupy a given position at a given time. De Broglie, Heisenberg, Schrödinger, Dirac and others all concur in this view that all apparently continuous phenomena are nothing but mean behaviors of inconceivably great numbers of elementary particles. When we have to deal with one isolated electron, for example, it is definitely impossible, according to the widely accepted Principle of Indeterminacy of Heisenberg, to measure accurately at one and the same time the position and the velocity of the particle. Either measurement alone may be made with very great precision; an increase in precision of one, however, must be compensated by a decrease in the precision of the other. Nor is this due to technological difficulties. The impossibility of determining both velocity and position is apparently due to the fact that they simply do not coexist in any real sense with reference to one electron. What is really measured—and what, therefore, alone

really exists in any scientific sense—is the probability that the electron will occupy a certain one of a number of possible positions.

In reply to the objection that statistical methods can not establish causal law, the writer will say that the only rigid "law" he can conceive is the law of probability.

Summary and Conclusions

(1) A need exists for a criterion of the exhaustiveness of factors employed in parole prediction. The only such criterion so far employed is that invoked by Laune.

(2) The optimum number of factors appears to be the maximum number containing the requisites: reliability, significance, orthogonality, stability.

(3) No factor should be employed, whose reliability coefficient is less than .50.

(4) No factor and no category should be employed unless it is significantly associated with outcome on parole. In the absence of time to perform more accurate calculations, the rule that in groups where the mean violation rate is in the neighborhood of 25% or of 75%, a factor should not be employed unless it differs from the mean by 5%, and in those groups whose mean rate is in the neighborhood of 50%, by at least 6%, is a satisfactory first approximation.

(5) Factors should be tested for inter-correlation, and in cases where such correlation exists efforts should be made to isolate the basic common factor. If this is impossible, that one alone of the group of correlated factors should be retained which shows the highest degree of association with outcome.

(6) Factors and categories should be tested for stability and all those categories whose index of instability is greater than .01 should be discarded. All new experience tables should be fully reported as soon as possible in order that the stability of categories may be better tested and in the interests of greater uniformity in the definition of categories.

(7) As between weighted and unweighted scoring methods the choice probably should lie with the system which uses the optimum number of weighted factors. Experimental tests to date are inconclusive. A total score based on the difference between success scores of favorable factors less those of unfavorable factors is suggested.

(8) Not only is the statistical approach valid in etiological studies, but, inasmuch as causality itself is a statistical concept, it constitutes the only approach ultimately possible.