SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Comparative Issues and Methods in Organizational Diagnosis.

5. TYPE OF REPORT & PERIOD COVERED
Technical Report.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
David G. Bowers
Alan S. Davenport
Gloria E. Wheeler

8. CONTRACT OR GRANT NUMBER(s)
N00014-77-C-0096

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Institute for Social Research
University of Michigan
Ann Arbor, Michigan

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Manpower Research & Development Program
Office of Naval Research
Arlington, VA

12. REPORT DATE
November 1977

13. NUMBER OF PAGES
96

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
93p.

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

DDC
RECEIVED
DEC 6 1977
F

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

adaptation	classification	knowledge base
adaptive plans	clinical judgment	Likert, R.
Bayes theorem	Decision Tree	model
beta distribution	Frequency-count-model	multiple discriminant function
CANOPUS	hierarchical grouping	normal distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

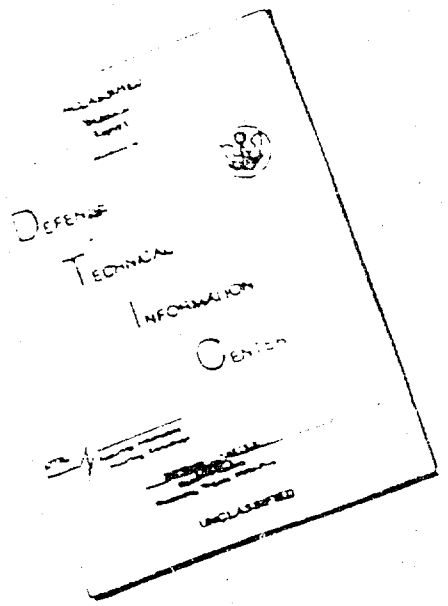
This report both introduces the topic of research on organizational diagnosis and discusses the methodological issues involved in it. It then describes the analyses to be reported in subsequent technical reports in the series. Four methods of diagnostic classification are to be examined using the Survey of Organizations data bank: distance function, multiple discriminant function, Bayesian, and decision tree.

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Succeeding technical reports in the series cover the following analyses:

1. Assignment of all 6,119 groups to one of the 17 types in the Bowers & Hausser typology by a distance function method. This procedure will constitute the "expert diagnosis" criterion.

2. Calculation of discriminant function weights from a developmental random half-sample of 3,000 (approximately) groups.

3. Development of likelihood functions to use in Bayes's Theorem, under the following conditions:

   (a) Independence of index measures:

      (i) Assuming an underlying normal distribution.

      (ii) Assuming an underlying beta distribution.

   (b) Non-independence of index measures.

4. Assignment by a decision tree procedure.

5. Comparison of the techniques in terms of accuracy-based criteria:

   (a) Proportion of correct classification for MDF, Bayes, and DT methods compared to a distance function criterion.

   (b) Reproduction of the typology.

   (c) Weighted dimensional distances.

   (d) Zero-one count.

6. Comparison of the techniques in terms of severity of misclassification, defined as costs associated with inappropriate treatment recommendations.

7. Comparison of the techniques in terms of information required, by (a) eliminating indexes, (b) collapsing indexes, and (c) sampling of respondents.

8. Comparison of the techniques in terms of ease of calculation.

Block 19 continued

organizational diagnosis
organizational functioning
organizational theory
profile scoring
psychiatry
standardized instrument
Survey of Organizations

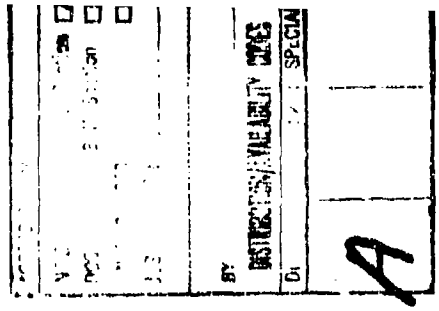


TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TECHNICAL REPORT

November, 1977

COMPARATIVE ISSUES AND METHODS IN ORGANIZATIONAL DIAGNOSIS

REPORT I

David G. Bowers
Alan S. Davenport
Gloria E. Wheeler

Institute for Social Research
University of Michigan
Ann Arbor, Michigan

This report was prepared under the Organizational Effectiveness Research Programs of the Office of Naval Research under Contract No. N00014-77-C-0096. Reproduction in whole or in part is permitted for any purpose of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

INTRODUCTION

No exact date can be designated as marking the birth of organizational development. Perhaps the late 1950's or early 1960's marked the first use of the specific term. What has taken place, therefore, has occurred within the last 15 years, years which have seen a substantial investment in the range of activities loosely representing this applied field. Although we lack exact dollar counts, a plausible estimate of the total funds invested in organizational development must run to hundreds of millions of dollars. By any standard, this is a large amount, one that no entity -- whether it be public or private -- may take lightly.

This same time period represents as well the first point at which it was conceptualized as organizational development, as opposed to management development or simply training. No exact definition has general currency, but the term is generally taken to refer collectively to an assortment of training or therapeutic interventions whose aim is presumed to be improvement of the organization and its members.

The reasons for organizational development's emergence at this particular time are not totally clear. Still, it no doubt ties to the series of traumatic national events which characterized the turbulent 1960's and early 1970's--war, assassination, inflation, energy crises, and political scandal. It was, in these short years, an intense period that called into question old ways of solving problems and old standards of behavior, while it called for new ones.

This applied profession's success in producing such constructive changes is another matter, however. Whatever its accomplishments may have been, they have been poorly documented, with the result that the base of

scientific knowledge in the field has been remarably thin. Kahn (1974) expressed it nicely:

> "A few theoretical propositions are repeated without additional data or development; a few bits of homey advice are reiterated without proof or disproof, and a few sturdy empirical generalizations are quoted with reverence but without refinement or explication." (p. 487)

He proceeded to cite a pointed example of the redundancy of the O.D. literature: of the 200 items contained in a prominent bibliography, only one-fourth included original quantitative data. One of the present authors (Bowers, 1976a) cited similar statistics: of the many titles listed for this field in the abstracts for these 15 years, only 18 studies other than doctoral dissertations can be found which contain real evidence, and a third of these are the work of three persons!

The same article catalogued a list of other shortcomings which organizational development must be concluded to have:

- Superficiality -- inadequate realism, inadequate relevance, and inadequate time available in the interventions undertaken.
- Commercialism - overenthusiasm, for fancy fees, and more promotion than careful design.
- Mistaken assumptions about the consultant's role - wasted and misplaced efforts stemming from faith in a "catalytic" stance which turns out to be content-less.

It is particularly to the last of these issues -- the consultant's role -- that the present report turns. An important part of the consultant's role is often presumed to be diagnosis -- translating a wide variety of symptoms into a coherent pattern that permits planning and carrying out appropriate remedial action. According to Lawrence & Lorsch (1969), the reasons for the importance of diagnosis in organizational development are many and persuasive:

(1) The client system may not be aware of the problem at all. For example, the difference between present effectiveness levels and unanticipated opportunities, rather than obvious difficulties, may be the "problem."

(2) The client system may not be aware of the real problem.

(3) A discrepancy between actual and desired outcomes does not explain and account for itself.

(4) Problem variance is likely to be multiply caused.

(5) Causes are likely to interrelate in complex ways.

(6) Causes are likely to differ greatly in potency, and what is desired is a designation of variables with leverage.

(7) Meaning can only be given to causal information by casting it into an appropriate configuration against a set of principles.

(8) What is required for action planning is an overall and integrated view, not a parochial one.

(9) Diagnosis, if done well, provides some insurance agains rushing into an inappropriate treatment that may prove damaging.

In contrast to this, the article cited earlier (Bowers, 1976) turned attention to assumptions concerning the consultant's diagnostic role in O.D. The points made there bear repeating.

- While a number of writers have attributed a diagnostic role to consultants, what goes unrecognized is that their diagnoses are often put to little other than heuristic use (that is, they are used merely to stimulate an interesting discussion).
- An unpublished study of consultants' diagnostic skills showed (a) inability to agree with diagnostic conclusions more formally obtained, and (b) more positive change occurring where consultants did relatively little diagnosing than where they did a great deal of it.
- Most consultants currently employ diagnostic methods which rely upon one observer--the consultant himself or herself--to obtain data. The N is restricted, not only in this fashion, but also by the fact that this consultant-observer is limited to a time-bound behavior sample.

These observations should not surprise us. Findings from the general field of assessment and classification have provided strong support to the position that statistical prediction is superior to non-statistical or judgmental methods (Cronbach, 1960). For example, in Meehl's (1954) major review of clinical versus statistical prediction, it was found that statistical prediction was equal to or superior to clinical prediction in 19 out of 20 cases.

Citing this body of accumulated evidence, Cronbach explores the reasons for perenially poor showings by (clinical) judges:

- Judges combine data by means of intuitive weightings which they have not checked.
- Judges casually change weights from one case to the next.
- Judges are unreliable, in the sense that the same case might not be judged the same way twice in succession.
- Judges have stereotypes and prejudices which affect their judgments.

His conclusions are the following:

> "What does this imply? It implies that counselors, personnel managers, and clinical psychologists should use formal statistical procedures wherever possible to find the best combining formula and the true expectancies for their own situation. They should then be extremely cautious in departing from the recommendations arrived at on the basis of the statistics..." (p. 348)

If this is the desirable state for organizational development as well, it is scarcely what in fact obtains. Levinson (1972, 1973), in his published remarks which led to the celebrated exchange with Burke (1973) and Sashkin (1973), stated that there is little resembling formal diagnosis in O.D. Consistent with Kahn's observations cited above, Levinson stated that the field is characterized by "ad hoc problem-solving efforts and a heavy emphasis on expedient techniques." Tichy (1975) does not reassure us when he finds, in his systematic empirical study, that change agents (consultants) seem to have limited diagnostic perspectives, that their

diagnostic frameworks are rather closely limited to their personal values and goals, and that the potential for intrusion of bias is not small.

Unfortunately, recommended alternatives are relatively scarce. Levinson's recommendations build upon a view and a method of organizational diagnosis that is an extension of the clinical case method. While large amounts of empirical data would be gathered, injecting a clinical judge between the data and the conclusion runs the risks listed above by Cronbach.

On the other hand, this is not the situation nor the age for "raw" empiricism. As the lengthy discussions nationally about discrimination in testing have revealed, in the interest of fairness and equal treatment, more must be taken into account in a decision process than any simple set of numbers, especially where connections between the numbers and real world events may not be obvious. In a similar vein, the sudden rise of the assessment center concept has shown that an appropriate criterion in this day and age (in employee selection, but by extension to the problem of treatment selection in O.D.) must include demonstrable connection between the measures used and the operations or functions performed in the real organization.

These facts lead us to the following preliminary conclusions, which form a starting point for the research to be undertaken in the present report and in the reports which follow it:

- The base of scientific knowledge which undergirds organizational development, while it is growing rapidly, is still remarkably small.

- Much of what is done is based upon consultants' predilections or fads, not upon solid reasons diagnostically generated.
- There is as yet little that could really be termed rigorous diagnosis practiced within the O.D. profession.
- Here, as elsewhere, statistical prediction is likely to prove far more accurate than clinical, or clinically mediated, prediction.
- Raw empiricism, in the form of predictors not obviously related to the processes and functions being diagnosed, no matter how seemingly accurate, are no longer societally acceptable. Prediction must be based upon measures derivable from solid scientific evidence about organizational functioning.

To understand what is or must be involved in diagnosis, we turn to a field which has practiced and taught diagnosis for years and decades, or even centuries: medicine. Ledley and Lusted (1959), in what must be counted as a seminal article, dealt at some length with the reasoning foundations of medical diagnosis. Table 1 presents a few of the principal points which they make, along with organizational diagnostic analogs. In the next sections we present a brief discussion of the content of each point.

TABLE 1

ISSUES IN DIAGNOSTIC REASONING

Medical Diagnostic Issues	Organizational Diagnostic Analogs
1. Symptom complexes (patterns of symptoms) are compared to disease complexes.	1. Patterns of actual organizational characteristics are compared to normative patterns of organizational characteristics.
2. Diagnoses are probability statements, not statements of certainty.	2. Diagnoses are probability statements, not statements of certainty.
3. Diagnosis aids the physician in selecting an optimum treatment under the ethical, social, economic, and moral constraints of our society.	3. Diagnosis permits the selection of an optimum treatment (intervention), given society's ethical, social, economic, and moral constraints.
4. The function of the knowledge base is to reduce the logical basis from all conceivable combinations of disease-symptom complexes to only those that actually occur.	4. The accumulated organizational knowledge base reduces available data to a manageable list of potential patterns of characteristics.
5. Maximizing the number of persons cured is equivalent to maximizing the probability that the individual patient will be cured.	5. Maximizing the number of units showing improvement is equivalent to maximizing the probability that an individual unit will show improvement.

Symptoms and Disorders

The total pool of available characteristics (of client units) is at any time limited to those which our knowledge base contains some information about and which our measurement methods are capable of measuring. All available characteristics are, at some level on their respective scales, potential symptoms. Whether they are, in fact, regarded as "symptoms" or not depends upon what past research and experience has found to be true -- that is, what has been added to the knowledge base.

What, then, are diseases, disorders, or states of organizational dysfunction? A disease is a hypothetical construct -- a theoretical term used for convenience purposes to refer to a whole chain of physical events which are hypothesized as having occurred. "Proof" that the hypothesized sequence has occurred (or is occurring) is obtained by some form of validation process. This validation can be concurrent or even retrospective: if little Johnny has influenza, he should display particular additional characteristics or should have displayed them within the last 24 to 48 hours. It can also take the form of construct validation, that is, of showing that only those observables that are hypothesized as going together in fact appear. Finally, the validation process can be predictive: we can wait to see whether subsequent, predicted signs of influenza appear in little Johnny's case. Throughout this sequence of comparisons, however, "influenza" is a hypothetical sequence of events which we presume to be able to see specific signs of at specific points in time. Its excellence as a classification category at any given point in the profession's development is entirely dependent upon the quality and completeness of the knowledge base from which we work, as it relates to the distinctions between this category and others.

What, then, determines what a disease is? It is the generalization and codification processes which past knowledge generators have gone through in integrating the findings from research and experience. Diagnostic procedures which rely upon "expert" assignment to diagnostic categories simply substitute the expert clinician for more public and replicatable listings. If the experts' procedures are unreliable, their classification is, as a criterion, worthless. If they are reliable and valid, it is a valuable aid -- a shortcut to employing the knowledge base directly and in its entirety.

Regardless of the way in which we mediate the process by which the knowledge base's contents get represented, the disease, disorder, or dysfunction is nothing other than a string of symptoms very much like those which we look at in any particular case. It is to this hypothetical symptom string or pattern that comparison is made in a diagnosis.

Diagnosis as Probability Statements

In organizational development and change, the diagnostic process follows essentially this same pattern. Symptoms are organizational characteristics which past research indicates go together to define some more general statement of organizational health or dysfunction. That our "diseases" do not have exotic names in Latin should not dismay us. Perhaps the absence of names at all is an advantage, in fact. Certainly there have been fewer years and resources available as yet for the codification of the knowledge base, and our professional schools teach us to be hesitant, cautious, and qualifying in our statements, rather than authoritative, definitive, and final. These are issues of style, however, rather than substance. The fact remains that there is an

existing knowledge base, comparison to which permits us to make a probability statement concerning any case at hand. Here, as in medicine, a diagnostic statement is a "best guess."

Relevance to Treatment Selection

The whole purpose of a diagnosis is to permit the selection of an optimum treatment or intervention. Here, as in the case of medicine, such choices are subject to social, ethical, economic, and moral constraints imposed by the society in which we live. Certain interventions may be socially unacceptable or even morally offensive. For example, intensely confrontational techniques are clearly unacceptable in many more traditional organizational settings, and under certain circumstances it is conceivable that top management team development training might generate an in-group clubbiness whose effects are racially discriminatory and therefore morally offensive. Other interventions, no matter how appropriate and promising, might be so expensive as to be prohibitive, while still others that would solve the problem might lead to violations of privacy and confidentiality which must be judged to be unethical.

However, within the limits which these constraints impose, the problem becomes one of selecting an optimum treatment from a pool of those available. What is optimum? Ledley and Lusted (1959) turn to value (decision) theory in an attempt to answer this question. Bowers and Hausser (1977) have shown how the organizational development problem can itself be cast into these same terms, and have presented empirical evidence about a limited number of intervention strategies.

A diagnostic procedure which clearly differentiates cases to which each of the known and available interventions are appropriate would obviously be superior to one which, in some measure or other, was unable to distinguish a condition calling for one intervention from a condition calling for another. At the most undesirable extreme would be a "diagnostic" procedure whose conclusions lead always to the same treatment or intervention, a condition which Levinson (1972) implied occurs in organizational development all too frequently.

Role of the Knowledge Base

Even with a relatively simple rating system of "Yes-No" or "Present-Absent," a list of N possible characteristics produces 2^N potential combinations. The number of potential "diseases" or dysfunctional states -- represented by the number of cells in an N-dimensional lattice -- is obviously unrealistically large. In any comprehensive scheme, all of the available units in the world probably would be insufficient to providing a single case per cell. The equally obvious conclusion is that most cells are empty, that they represent nonexistent disorders, and that only a relative few comprise the set of "real" possibilities. It is the task of the knowledge base to provide us with current, accurate information about what these possibilities are.

Much the same point is made in the theory of adaptation in natural and artificial systems (Holland, 1975). Combinations of characteristics rapidly generate astronomical numbers of possibly adaptive structures. If the organism or system were to choose an enumerative adaptive plan -- simply running down the list randomly

until it found the one that worked -- adaptation would rapidly become impossible. As the writer just cited indicates, given even the fastest computers in existence, it would require "a time vastly exceeding the age of the universe to test 10^{100} structures." Instead, adaptive plans to be feasible must be robust -- that is, they must be efficient over the range of situations which will be encountered. One general requirement, therefore, is that an adaptive plan must retain advances already made, as well as parts of the history of what has occurred. This information, of course, is what constitutes the knowledge base in any diagnostic system.

Improvement Probabilities and the Single Case

At first blush, the statement seems unfeeling or insensitive that we maximize the probability of any individual unit's showing improvement when we apply to it a strategy shown to maximize the number of units showing improvement. Organizational development is, after all, a human practice profession, and it seems impossible to ignore facts obviously at hand (within range of our personal observations, for example).

Nevertheless, observations based upon an N of one (observer), collected under atypical conditions and within nonrepresentatively short time frames, are no more reliable and accurate when taken singly than they would be if used en masse for large numbers of cases.

This issue was touched upon by one of the present writers in an earlier report: "Even the most accurate diagnosis may suffer from mid-stream or horseback revisions made by the consultant as he approaches its use. Basically, any data collection and analysis method treats with some

degree of care and accuracy a portion, but not all, of the behaviors, events, and issues in the life space of the client system. Some portion is unique to that system, or to any group within it, or will have been excluded from the array of information categories designed in the diagnostic process at its inception. As the consultant approaches a particular unit or group of the client system, he will necessarily see other aspects of what he feels are its functioning not represented in the diagnosis which he has in hand. Since he is dealing with a real client, in a real world situation, the temptation is well nigh irresistable to revise the diagnosis on the basis of his current observation. Yet, he is one observer observing at best a limited and time-bound behavior sample. To the extent that he makes such revisions he therefore very likely reduces both the reliability and the validity of the diagnosis with which he works. Said otherwise, he approaches each group, or each setting, as a unique instance with live people and real problems. Yet in many ways the diagnosis and treatment problem in organizational development is a "large N" problem. Were he to work on the basis of the diagnostic data provided to him and that alone, given that it is reliable and valid, he would, across a large number of cases, succeed in a high portion (assuming that the diagnostic and prescription processes are themselves high in quality, reliable, and valid). Yet he does not ordinarily approach his role with that degree of objective detachment, and each time that he yields to the temptation to revise on the basis of "current reality" he submits himself to a situation in which his action steps are based on less than acceptably reliable and valid data." (Bowers, 1974)

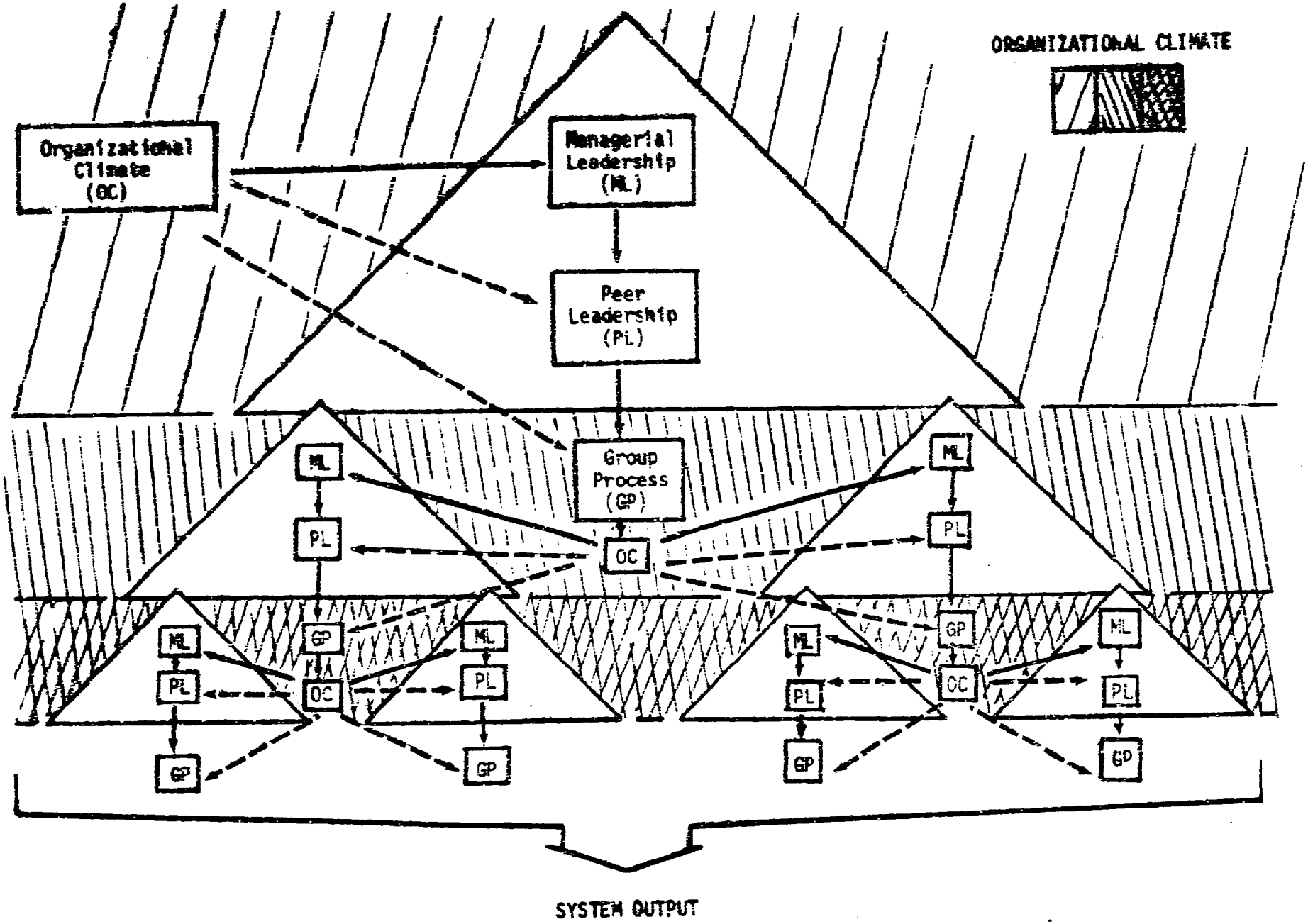
Toward Relevant Research

Clearly, therefore, any attempt to develop and test more rigorous diagnostic procedures in organizational development should be based upon a model containing principles of organizational functioning. In other words, it should be theoretically anchored to a conceptual statement that is itself both organizational in content and comprehensive in scope. While the literature on organizational management is ripe with theoretical statements, most of them do not meet criteria of acceptability for our present purposes. Many must be dismissed as less comprehensive than is necessary for the present problem: that is, they are elegant treatments of an isolated issue such as job design or individual satisfaction or leadership, but they ignore other areas. Others may be rejected because, although they encompass most of the domain, they are lacking in adequate empirical underpinnings. However, one theoretical statement which does appear to meet the criteria just outlined is the Likert meta-theoretical paradigm. (Likert, 1961, 1967, 1976; Bowers, 1976) It is this theoretical statement which underlies the measures collected in the data bank to be used in the research launched by this project.

Most recent evidence suggests that this paradigm assumes the form taken in Figure 1 (Bowers & Franklin, 1977). As a set of principles, this paradigm would appear to satisfy the criteria of comprehensives and evidential validity. (Bowers & Franklin, 1977; Likert, 1977). It is operationalized here in the form of the Survey of Organizations, a machine-scored standardized questionnaire which has been used in various editions since 1966 to collect organizational survey data for assessment, feedback, and benchmark purposes (Taylor & Bowers, 1972). Portions of these

Figure 1

ORGANIZATIONAL FUNCTIONING



banked data have been used in earlier research efforts related to organizational development. In this regard, a method of diagnostic classification was previously developed and preliminarily tested. Termed CANOPUS, it contains a software package designed to generate a diagnostic statement for groups and pyramids of groups comprising organizations (Bowers, 1974). This classification method is based upon a typology of work groups developed in the course of prior research. The technique used for the development of the typology was profile analysis, in which one arrives at a clustering of work groups. The profile consisted of a group's scores on the SOO indexes and as a profile reflected three basic kinds of information: level, dispersion, and shape. Level was the mean score of the work group over the indexes in the profile; dispersion reflected how widely scores in the profile diverged from the average; and shape concerned the profile's high and low points.

A measure of profile similarity that takes shape, level and dispersion into account is the distance measure. If one considers a person (or group) as a point in a multidimensional space in which each dimension represents a variable or index, then the distance between two points, that is, persons (or groups), can be computed using the generalized Pythagorean theorem. The distances can then be examined to determine which groups cluster together in that multidimensional space.

The clustering technique, called hierarchical grouping, uses this distance measure as a measure of profile similarity. Computer software is available for this technique in the program, HGROUP (Veldman, 1967.)

This program begins by considering each original object, in this case a work group, of these to be clustered, as a cluster. These N clusters are then reduced in number by a series of step decisions until all N objects have been classified into one or the other of two clusters. At each step the number of clusters is reduced by one through combining some pair of clusters. The particular pair to be combined at any step is determined by the computer's examining all the available combinations and choosing the one which minimally increases the total variance within clusters. It is this latter minimizing function that utilizes the distance notion. The total variance within clusters is a measure of the closeness of the points in multivariate space in clusters already chosen. A substantial increase in this variance, which the HGROUP program labels an error term, indicates that the previous number of clusters is probably optimal for the original set of objects or work groups. The program provides an identification of those groups contained in each cluster so that further analyses can be conducted on phenomena within clusters.

The HGROUP program was applied to three random subsamples drawn from the data bank (Hausser & Bowers, 1977.) When the three sets of data were considered jointly, a total of 17 distinct profiles emerged. In many ways, this software system would appear to meet generally the requirements listed:

- It compares data to appropriate norms.
- Problems once identified are prioritized in terms of their potential impact upon outcomes.
- It seeks causes for observed conditions among situational, information, skill and values conflict predictors, employing a distance statistic.

- It selects a broad set of potentially appropriate action steps from an array of possibilities.
- It converts the whole and its parts into a readable narrative by computerized text-writing.

Still, the method is based upon the measures, and those measures derive from the theoretical paradigm previously cited. While attractive, it is but one of several statements that might have formed the basis for operations and measures. Clearly some difference among theorists is to be expected. The domain is sliced differently, and the terms applied to collections or clusters of behaviors and processes will vary substantially. However, if the fundamental, general algorithm is the same, we can at least be somewhat reassured that subsequent work will not be unacceptably parochial.

In an effort to address this question, the writings of nearly 30 prominent persons in the field (including Likert) were examined. While no effort was made to consider all possible positions taken by every conceivable writer, an effort was made toward at least reasonable comprehensiveness. All of those considered were concerned in one way or another with organizations, and all had demonstrable action interests. Some were discarded after a brief scrutiny for the reason that their interest appeared to be non-organizational, that is, that, for example, the outcome valued was personal growth or individual adjustment, not organizational effectiveness. Others were discarded because their writings were restricted almost totally to the operation of a particular technique in a limited environment (management by objectives, for example). In the end, more careful consideration was given to the writings of 11 theorists or pairs of theorists. (See Table 2.)

Table 2

Theorist Examined and Principal Writings

Theorist	Principal Writings
Argyris, C.	Personality and organization. New York: Harper & Bros., 1957 Intervention theory and method: a behavioral science view. Reading, Mass.: Addison-Wesley, 1970.
Cyert, R.M. & March, J.G.	A behavioral theory of the firm. Englewood Cliffs, N.J.: Prentice-Hall, 1963.
Davis, S.A.	An organic problem-solving method of organizational change. J. of Appl. Beh. Sci., 1967, 3, 3-21.
Fiedler, F.E.	A theory of leadership effectiveness. New York: McGraw-Hill Book Co., 1967.
Herzberg, F.	One more time: how do you motivate employees? Harvard Business Review, 1968, 46, 53-62.
Kilmann, R.H.	Social systems design: normative theory and MAPS design technology. New York: North Holland Publishing Co., 1977.
Lawler, E.E.	Motivation in work organizations. Monterey, Calif.: Brooks/Cole Publishing Co., 1973.

Table 2 continued

Lawrence, P.R. & Lorsch, J.W.	Developing organizations: diagnosis and action Reading, Mass.: Addison-Wesley Publishing Co., 1969.
Likert, R.	New patterns of management. New York: McGraw-Hill Book Co., 1961. The human organization. New York: McGraw-Hill Book Co., 1967. New ways of managing conflict, New York: McGraw-Hill Book Co., 1976.
McGregor, D.	The human side of enterprise. New York: McGraw-Hill Book Co., 1960. The professional manager. New York: McGraw-Hill Book Co., 1967.
Steele, F.I.	Physical settings and organizational development. Reading, Mass.: Addison-Wesley Publishing Co., 19

At first blush, the positions represented by these persons would appear to be diverse. Herzberg, after all, is primarily oriented toward job enrichment, while McGregor's position was one focused around the behavioral effect of personal values. However, a closer look reveals that, with minor variations, the same general algorithm underlies most, if not all. While it varies greatly in form, and considerably in the exact items noted, recorded, or counted, it assumes that something external to the organism (group, group member, or organizational unit), that is, something in its environment, combines or interacts with something internal to the organism. This process leads to states of internal functioning on the part of the organism that in turn result in effectiveness. Stated at a highly general level, it can be seen as an expression of the old Lewinian equation, $B = f(P,E)$.

Some difference among theorists obviously occurs around the issue of aggregation of individuals into collectivities such as groups. Some (as, for example, Lawler or Herzberg) treat individuals separately and integrate afterward, sometimes by an implicit summation. Others (for example, Lawrence & Lorsch or Likert) aggregate first and deal principally with group-level processes and results. In the light of the similarity of the general algorithm, however, this difference seems comparatively minor.

Let us consider in greater detail each of the components, environment and person. Turning first to the <u>environment</u> segment, words, terms and phrases vary widely, but nearly all of the theorists concerned seem to focus upon the following set of interrelated issues:

- information flows, processes, and patterns
- motivational conditions, structures, and systems
- task configurations, structures, or flows
- norms, values, attitudes, and beliefs stemming from superordinate governance systems
- distribution of power, particularly for resource allocation
- technical or physical conditions, or configurations of objects

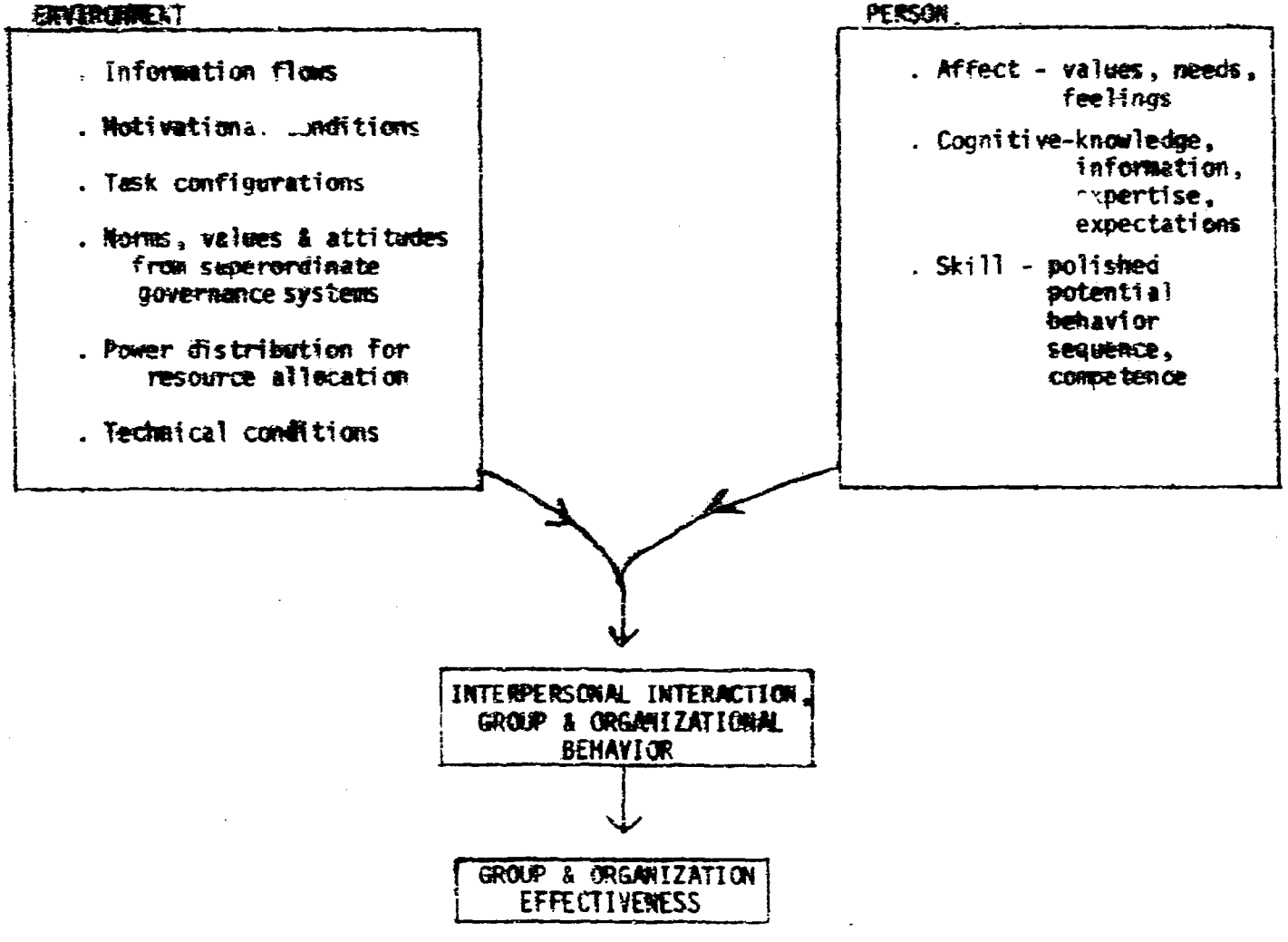
The person portion of the algorithm is variously represented in the different theoretical treatments, but appears to deal in the following:

- an affective component, in the form of values, needs, or personality orientation
- a cognitive component, in the form of knowledge, ability, information, expertise, and expectations.
- a behavioral potential component, in the form of competence and skills.

In some instances, important modifications of one or more of these are assumed to be represented by such personal background descriptors as age or education. Even where this is so, however, it is clear that it is the affective, cognitive, and behavioral implications, rather than sheer demographic facts, that are held to be important.

Thus the algorithm might be restated in somewhat more elaborate fashion as it appears in Figure 2. The problem, of course, becomes more complicated when extended to those groups of groups called whole systems. Outputs from one group become environmental situations forming inputs to other groups. (Bowers & Spencer, 1977).

Figure 2


ENVIRONMENT
. Information flows
. Motivational conditions
. Task configurations
. Norms, values & attitudes from superordinate governance systems
. Power distribution for resource allocation
. Technical conditions
PERSON
. Affect - values, needs, feelings
. Cognitive-knowledge, information, expertise, expectations
. Skill - polished potential behavior sequence, competence
INTERPERSONAL INTERACTION, GROUP & ORGANIZATIONAL BEHAVIOR
GROUP & ORGANIZATION EFFECTIVENESS

Nevertheless, we come away from this examination of the field reasonably confident that a common algorithm underlies most of the major works in it. It leaves us reassured that adherence to an alternative formulation from this list, if pursued to its most basic form, would not result in an utterly different diagnostic scheme. Terms might be different, and the operations employed by each writer to measure particular sets of variables might vary widely, but the rationale and the set of primitive constructs would be very much the same.

A Stock-Taking and Some Implications

Against the expressed need for improved methods for diagnosis in organizational development we can array the following major points from the preceding discussion. An adequate diagnostic procedure necessitates:

(1) A theoretical model which is acceptably comprehensive, which shares the same general algorithm present in the array of principal alternative formulations.

(2) A bank of data, collected by a standardized instrument in a wide variety of organizational settings, both military and civilian.

(3) A recognition that accuracy in diagnosis will, here as elsewhere, very likely be enhanced by statistical operations rather than clinical judgment.

(4) An acknowledgment of societal requirements rejecting "raw" empiricism in favor of statistical procedures and measures which are content valid.

Considering the magnitude of the problem, the size of data sets, and the turnaround time requirements present in most organization development situations, yet another requirement would appear to be present: that whatever operations result be computer-assisted. In this area, one can profit from the experience of another practice-oriented profession whose researchers have explored computerized diagnostic procedures, psychiatry.

There, as elsewhere, substantial difference of opinion exists concerning the best method of evaluating the importance of symptoms. Two general types of models relying upon probability statistics have been proposed:

(1) A discriminant function model, in which each symptom is given an empirically derived weight, and an artificial measure is then obtained as the sum of the weighted values (Crooks, Murray & Wayne, 1959).

(2) A Bayesian or frequency-count model, in which the relative frequency of occurrence of each possible symptom-disease pattern is considered (Ledley & Lusted, 1959).

To these has been added a third method:

(1) Treats the issue as a decision-tree problem, thus relying maximally upon excellence of the knowledge base and not at all upon probabilities in a developmental sample (Spitzer & Endicott, 1968).

The Three Methods in Detail

Multiple Discriminant Method: This method uses a large developmental sample of previously diagnosed cases and is especially suited for quantitative data. It takes account of correlations among the measures and finds for each clinical group in the developmental sample a set of weights derived so as to maximize each subject's likelihood of being assigned to the group from which he came. In subsequent use, each new case is assigned to the group for which the composite score represents the least distance. The discriminant function model uses a diagnostic equation of the following type:

$$Y = aX_1 + bX_2 + cX_3 + \ldots + pX_n$$

where the coefficients a, b, c, ... , are weights proportional to the magnitude of the relationship between the particular symptom and the disease process, reduced according to the relationship of this symptom to the others represented in the equation, and X_1, X_2, etc. are the observed symptom values.

The arguments in favor of a multiple discriminant model are at least threefold:

(1) It better replicates the thought process employed by the human diagnostician, who does not treat each symptom in present/absent fashion, but rather attaches greater or less weight to each symptom according to his past experience, (i.e., his own version of the knowledge base.)

(2) Symptoms which correlate highly with the presence/absence of a disease are given more weight than those that have shown little or no correlation to its presence/absence.

(3) Appropriate weighting also depends upon a symptom's correlation with other symptoms. If the overlap is high, then one would weight the second symptom much lower than would be the case if the two symptoms have little relationship to each other.

At least two objections have been raised to this method:

(1) It relies upon the accumulation of a large developmental sample of cases, which is difficult, expensive, and in most instances unlikely.

(2) It capitalizes upon accidental features of the developmental sample and thus gives an inflated estimate of its accuracy. If the validation sample comes from a somewhat different population, the drop in efficacy is even greater.

Still, the method has been explored and developed (Rao & Slater, 1949; Melrose, et al., 1970; Sletten, et al., 1971). It is closely aligned, although not identical, to the method employed by our own effort at computerized organizational diagnsos, which employs the distance statistic (D) in a hierarchical grouping procedure.

Bayes Classification Method: This method also uses a large developmental sample of previously diagnosed cases to determine, for each diagnostic category, the relative frequency of each possible profile of scores. Its fundamental diagnostic formula is:

$$E \longrightarrow (G \longrightarrow f)$$

where E is scientific knowledge; G is a complex of symptoms; and f is a complex of diseases. It may be read, "Existing scientific knowledge E implies that, if symptom complex G is in evidence, the patient very probably has disease f."

The knowledge base is thus stated as a series of conditional probabilities, of the form P (G | f), i.e., the probability of a patient's displaying the symptoms, given that the disease is present. It may be "converted" to the more diagnostically useful form P (f | G) -- the probability of a patient's having the disease, given that the symptoms are present -- by Bayes' formula. By this formula, the probability of a patient's having a particular disease, given that he has a particular complex of symptoms, is equal to:

- the probability of the disease's occurring in the population at all, multiplied by
- the probability that a patient will display this particular complex of symptoms, given that the disease is present, and this result divided by
- the summation over all disease complexes of the probability of the disease complex multiplied by the probability that a patient will display this particular complex of symptoms, given that (each individual) disease complex is present.

Thus knowledge base statistics and symptomatic information from the current case can be combined to establish the needed diagnostic probability. A subject with a given profile is then assigned to the group for which his profile is relatively the most frequent.

The method is especially suited for categorical or nominal data expressed in mutually exclusive categories. It can also be applied to numeric data by grouping scores into intervals and treating each interval as a qualitative category, ignoring both its original quantitiative value and its ordinal position.

The principal argument raised in favor of a Bayesian approach to computerized diagnosis appears to be that it also is claimed to model the human judgment process by which symptoms are converted into a diagnostic statement (that is, that the physician, for example, employs a conditional probability judgment process in arriving at a diagnosis.)

The objections are a bit more extensive:

(1) It is difficult, if not impossible, in diagnostic work to satisfy the conditional independence requirement (the requirement that the probability of finding one particular symptom given that the disease is present, is unaffected by the presence or absence of any other symptom.)

(2) As in the other statistical method, it requires the accumulation of a large developmental sample of appropriate form and content.

(3) The necessary assumption that the diseases are mutually exclusive may not hold.

(4) As in the other statistical method, it capitalizes upon accidental features of the developmental sample.

Regardless of these objections, the method has been explored, developed, and experimentally implemented in psychiatry (Birnbaum & Maxwell, 1960; Overall & Gorham, 1963; Smith, 1966).

Decision-Tree Method: The logical decision tree method starts with a series of questions, each of which is treated as true or false. The true or false response to each question rules out one or more diagnostic possibilities and determines which question is to be answered next. The questions may specify the presence of a single symptom, the existence of

a numeric score in a certain range, or a complex pattern of symptoms and scores. The method ordinarily results in the subject's being assigned to a single diagnostic category, with no quantitative measure of similarity to either that category or other categories.

The arguments in favor of this method are given by at least one proponent (Spitzer, et al., 1974) as the following:

(1) It is independent of any specific body of data; that is, it does not require a large developmental sample.

(2) It is not constructed so as to be optimal for any one population and for this reason "travels well" from one setting to another.

(3) As in the case of each of the other two methods, it is thought to represent optimally the thought processes of the human diagnostician.

The objections are the following:

(1) It is quite dependent upon the accuracy of the theory which underlies the decision tree and is therefore ultimately as dependent as the other methods upon past data accumulations, their care and form.

(2) Its generalizability may be more apparent than real.

(3) Its assumption that the diseases are mutually exclusive may not hold.

Approaches following this method have been developed and applied by several investigators, with a variety of results (Spitzer & Endicott, 1968; Wing, 1970).

Comparison of the Three Methods

Several efforts have been undertaken in psychiatry to compare two or more of these methods empirically. The results are best described as decidedly unclear. Overall and Hollister (1964) conducted one such comparison, but, unfortunately, the rules used by their computer programs were obtained from diagnostic stereotypes provided by experts, rather than from observed characteristics of actual cases.

Melrose, et al. (1970) compared a multiple discriminant with a decision-tree approach and found that: (a) on single assignments the decision-tree approach showed a greater degree of agreement with an expert judgment criterion; (b) if first, second, or third possible assignments were allowed, the multiple discriminant method showed a greater degree of agreement than did the decision-tree; and (c) in any event, each method performed better for certain diagnostic categories.

Finally, Fleiss, et al. (1972) compared all three methods and found none of the three to be clearly superior to the other two. Again, however, the criterion was agreement with expert diagnoses, a criterion whose unreliability the authors duly note.

A Dilemma and Some Issues

The work from psychiatry, just cited, contains a dilemma whose existence questions the whole body of findings and whose resolution might be seen as rendering the whole exercise rather trivial. Elegant, replicatable, and readily transportable methods are designed and tested against a criterion of "judgment by expert clinicians." Yet here, as elsewhere, the

Cronbach warnings apply: expert clinical judgment is notoriously unreliable. What has been developed, therefore, are three elegant ways of replicating an unreliable procedure. On the other hand, had a reliable, replicatable, transportable procedure existed for use as a criterion, it would no doubt have been more sensible to employ it as the diagnostic method, rather than as a criterion for other methods.

Several implications stem from this observation. First, where in psychiatry expert clinical judgment is an unreliable classification method and criterion, in the present instance we do possess a reliable, verifiable procedure, one based upon the distance statistic. Obviously, if accuracy is interpreted in distance terms, no alternative procedure can be as accurate as that self same distance statistic. If, therefore, this were the sole or major issue to be researched, the project would end immediately.

It is not, however; other issues of equal or greater importance occur, the answers to which are in no way obvious. These issues, to be explored in the sections which follow, include: (a) proportion of correct classification, (b) typology reproduction, (c) weighted dimensional distances, (d) zero-one counts, and (e) information reduction. Separate from this is the whole issue of a decision-tree (with or without Bayesian inputs), and the non-substantive matters of efficiency, cost, and ease. These latter issues also are dealt with.

BASIC PREPARATION OF THE DATA

The purpose of this section is twofold: to describe the data set to be used for the remainder of this study; and to describe the four classification techniques (distance statistic, multiple discriminant function, Bayes, and decision tree) as they will be used in the research to be subsequently reported.

Description of the Data Set

The existing national (civilian) normative file of the Survey of Organizations (SOO) contains 5,599 groups. It represents the total body of data collected since 1966 from some 37,098 persons in a broad segment of the civilian industrial population. As such, it represents many different industries, functions, and hierarchical levels.

Available also are data from two independent military samples. The first of these contains more than 200 usable groups of Navymen from whom questionnaire data on SOO indexes were collected in late 1972 and early 1973. The second contains 320 groups of Army soldiers from whom data were collected in late 1974 and early 1975. In each of these instances, in order to satisfy the need for intact units, it was decided to collect data from all members of a selected number of organizational subunits or "modules." These modules consisted of a pyramid of work groups three echelons, or tiers, tall. Thus data were collected from all members of the three organizational levels immediately below a designated "module head." Modules themselves were selected by what amounts to a

stratified random sampling procedures. Methods are spelled out in greater detail in two technical reports (Michaelsen, 1973; Spencer, 1975.)

Taken together, these various data sets comprise a sample of 6,119 groups, to be employed in the main analyses. For other analyses, performance measures may also be employed for a subset containing 940 groups for which measures of efficiency and/or attendance are also available. The measures themselves have been examined extensively in the course of another project and their properties reported (Pecorella & Bowers, 1976a; 1976b.) They have been converted to standardized score form to attain inter-organizational comparability.

From the onset of the project to the present time, these various data sets have been reformated so that all share a common format. All data have been entered into a single large file with merged superordinate values (those from the group immediately above), plus with merged values for a second wave of survey data where such a second wave exists.

Measures Used

The Survey of Organizations contains in its 1974 edition 16 standard indexes. Two of these, because they have not been universally used since the start of the data bank, will be dropped. The 14 which remain will form the survey index measures to be used in the present study:

Organizational Climate

Decision Making Practices -- the manner in which decisions are made in the system: whether they are made effectively, made at the right level, and based upon all of the available information (4 item index).

Communication Flow -- the extent to which information flows freely in all directions (upward, downward, and laterally) through the organization (3 item index).

Motivational Conditions -- the extent to which conditions (people, policies, and procedures) in the organization encourage or discourage effective work (3 item index).

Human Resources Primacy -- the extent to which the climate, as reflected in the organization's practices, is one which asserts that people are among the organization's most important assets (3 item index).

Lower Level Influence -- the extent to which non-supervisory personnel and first-line supervisors influence the course of events in their work areas (2 item index).

Supervisory Leadership

Supervisory Support -- the behavior of a supervisor toward a subordinate which serves to increase the subordinate's feeling of personal worth (3 item index)

Supervisory Team Building -- behavior which encourages subordinates to develop mutually satisfying interpersonal relationships (2 item index).

Supervisory Goal Emphasis -- behavior which generates enthusiasm (not pressure) for achieving excellent performance levels (2 item index)

Supervisory Work Facilitation -- behavior on the part of supervisors which removes obstacles which hinder successful task completion, or positively, which provides the means necessary for successful performance (3 item index)

Peer Leadership

Peer Support -- behavior of subordinates, directed toward one another, which enhances each member's feeling of personal worth (3 item index).

Peer Team Building -- behavior of subordinates toward one another which encourages the development of close, cooperative working relationships (3 item index).

Peer Goal Emphasis -- behavior on the part of subordinates which stimulates enthusiasm for doing a good job (2 item index).

Peer Work Facilitation -- behavior which removes roadblocks to doing a good job (3 item index).

Satisfaction -- a measure of general satisfaction made up of items tapping satisfaction with pay, with the supervisor, with co-workers (peers), with the organization, with advancement opportunities, and with the job itself (7 item index).

The typology of work groups to be used in this study is reported in Bowers and Hausser (1977), contains 17 types, and is based on the indexes of the Survey of Organizations. The resulting types have different profiles across the indexes, with the patterns of these profiles being quite distinct. In the present project, for purposes of evaluating the different classification procedures, all work groups will first be placed into one of the 17 types by each of four methods: distance function, decision tree, multiple discriminant function, and Bayes. A description of each of these techniques and their implementation in this study is given below.

Distance Function Method

The HGROUP program used in the original generation of the typology, described earlier, uses a generalized distance function based on the error sum of squares -- the squared deviations from group means. In the case of the classification system applied to new or "incoming" groups, this present method calculates a similar distance statistic between the profile of indexes for the group to be classified and the profiles (mean index values) for each of the 17 standard types. The group is then assigned to the type for which the distance value is smallest.

All 6,100 (approximately) work groups will be so classified by the distance function method. This classification will serve as the "expert" or correct classification in this study. From this assignment it will be possible to calculate the vector of means and the variance-covariance matrices for each type, as well as estimate the distribution of work group types.

Decision Tree Method

A decision tree is effectively, a sequenced series of questions, each with a limited set of alternatives, which, when one is selected, leads to the next branch of the tree and to the next question. A simple example is given in Figure 3. This example contains three questions (A, B, & C), the first two with two alternatives each, the third with three. The result is twelve end states (not all of which need be distinct; only the routes getting there need be distinct.) In applications to classification, the end states represent the categories of the typology.

The development of a decision tree is not a data based procedure, but rather, a theoretical one. For the purposes of diagnosing work groups into the existing seventeen-class typology, a decision tree whose questions will relate to relative performance on the 14 indexes of the SOO is presently being developed. The actual format of this decision tree will be presented in a separate technical report.

The decision tree algorithm will be used to classify all (approximately) 6,100 work groups in the data set.

Multiple Discriminant Function Method

The use of multiple discriminant functions to assign units into classes within a typology is a standard procedure. Essentially, the process is to find linear (or quadratic) functions of the predictor variables which maximally differentiate among the groups of the typology. With more classes in our typology than predictor variables, the maximum number of discriminant functions is equal to the number of variables,

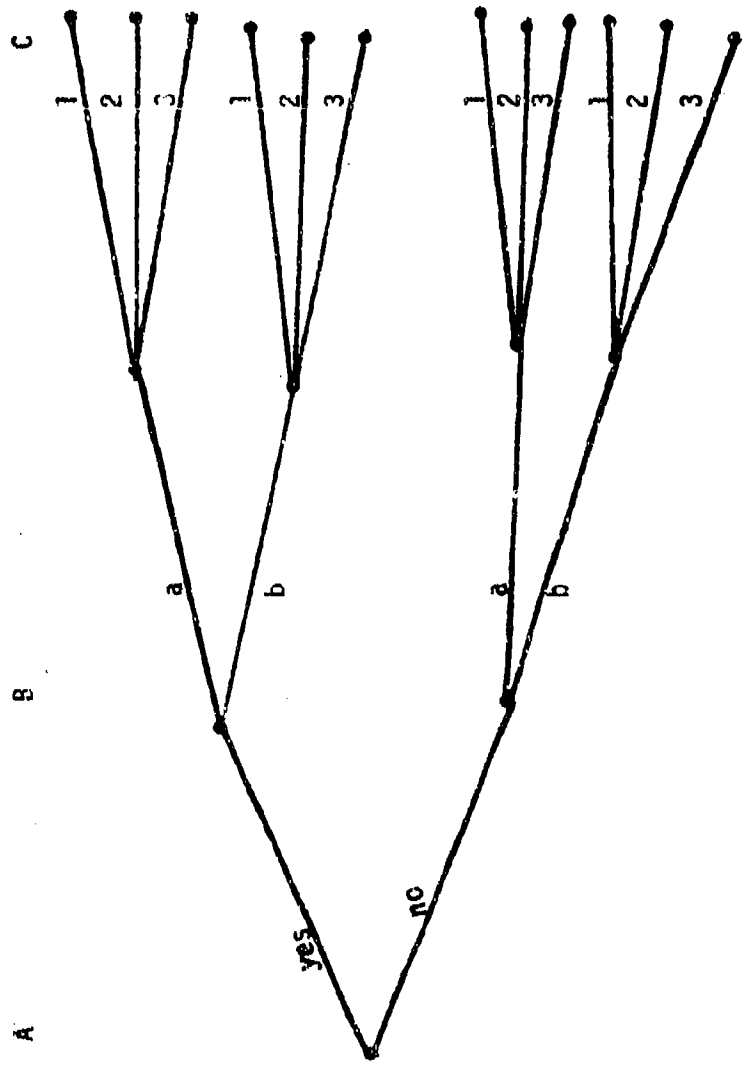


Figure 3

Question

say r. While typically almost all of the explainable variance is accounted for by the first two discriminant functions, it is possible to utilize all r functions. The primary value in the use of discriminant functions is that they are orthogonal.

Notationally, the ith discriminant function is given by:

$$Y_i = \sum_{j=1}^{r} a_{ij} X_j$$

where $a_{i1}, a_{i2}, \ldots, a_{ir}$ are the weights and $X_1, X_2, \ldots, X_r$ are the predicter variables. Thus the r discriminant functions are given by $Y_1, Y_2, \ldots, Y_r$. The procedure for finding the weights a_{ij} results in Y_1 explaining the most variance, Y_2 the second largest, etc. In particular, the weights a_{ij} are chosen to maximize

$$\frac{\text{variance between means on } Y_i}{\text{variance within groups on } Y_i}.$$

The weights a_{ij} are calculated from a developmental sample. Once they are calculated, the next step is to calculate the vector of Y values for each work group, say $\underline{y}$. For each class within the typology, the means of the $\underline{y}$ vectors, say $\underline{\mu}$, and the variance-covariance matrices, are calculated. Finally, with the above information, one is ready to begin the classification process.

Under certain distributional assumptions, including equal variance-covariance matrices across types, the assignment of a new work group requires the calculation of the vector of discriminant functions, $\underline{y}$, and the Mahalanobis distance between $\underline{y}$ and the center of each of the types.

The actual assignment is to the type to which $\underline{x}$ is the closest. If one is not able to assume the common variance-covariance matrices, the assignment procedure, while still possible, is more involved. It effectively is a maximum likelihood procedure which then classifies a work group into the type which has the largest probability.

In this study, the actual discriminant functions will be calculated from a developmental sample of N=3,000 (approximately), which represents half of the total data set. (The developmental sample will be selected randomly from the entire set.) The use of a subset as a developmental sample will permit the assignment of both the developmental sample and the remainder of the data set as a check on the generalizability of the procedure across similar data sets. The actual computer operations to be performed are those in the DISCRIMINANT routine within MIDAS, the statistical package available at the University of Michigan (Fox & Guire, 1976.) Included in the output is a test of the homogeneity of variance-covariance matrices and the distances between the means of the types.

Bayesian Method

As earlier sections of this report have indicated, the role of the organizational development specialist can be linked to that of the physician. In both instances, the specialist is confronted with an entity (a person or an organization) which exists in some "disease" state. The specialist must diagnose the disease and prescribe treatment.

Ledley and Lusted (1959) described the use of probabilities in medical diagnosis. The approach that they used, which makes use of Bayes's Theorem, has been called a "Bayesian" approach to diagnosis, because it uses Bayes's Theorem as the model for developing final probabilities for each possible disease state. It is also called a relative frequency approach because it uses relative frequencies derived from historical data to develop the probabilities that enter Bayes's Theorem.

Historically, probabilities have been defined in terms of relative frequencies. For instance, the probability of obtaining tails in a fair coin toss is .5. This probability comes from the fact that if one were to toss a coin or coins a large number of times, the relative frequency of tails would be about .5. Intuitively, this definition of probability makes sense, at least when the events in question are truly repeatable.

There is another school of thought, however, that finds the relative frequency notion unduly restricting. Consider a unique event, such as the election of the President of the United States in 1980. Such an event occurs only once in history, and from a relative frequency point of view, it would be meaningless to talk about probabilities, e.g., the probability that a particular individual will win that election. The probability of the event is either zero or one, depending on whether it occurs or does not occur. On the other hand, it is perfectly clear that oddsmakers and others involved in betting are not restricted to relative frequency notions. Where relative frequency data are available, a perfectly rational oddsmaker would give bettors a set of odds that exactly match the relative frequencies (ignoring questions of commissions, "house cut," etc.).

Where relative frequency data are not available, however, the oddsmaker relies on a subjective probability. That subjective probability is a number between zero and one which represents the extent to which the person believes that a certain statement is true.

The Bayesian school of statistics defines probability in the above sense. To a Bayesian statistician, all probabilities are subjective probabilities. They represent degrees of belief rather than relative frequencies. Subjective probabilities should conform to all rules of probability theory just as do probabilities defined in some other way. The Bayesian approach allows the statistician to consider a broader amount and type of information, incorporating it into his calculations, than does a classical approach.

The Bayesian approach proposed in this section is similar to that which has been used in medical diagnosis, since it uses Bayes's Theorem and also relies on historical data. It goes beyond historical data, however, by using subjective probabilities rather than simple relative frequencies. The reasons for the approach will become more apparent shortly. In the meantime, however, a brief introduction to Bayes's Theorem is in order.

Bayes's Theorem

The diagnosis situation is characterized by uncertainty. The diagnostician is faced with a set of mutually exclusive and exhaustive states (diseases) into one of which the organization must belong. The diagnostician's job is to identify correctly the category that the organization

belongs in. The organization brings to the situation certain characteristics, akin to symptoms in the medical context, that provide the diagnostician with information about which disease states are likely to occur.

Bayes's Theorem is a relatively simple equation that defines the probabilistic relationship between symptom combinations and the resulting disease states. For the moment, consider probabilities as relative frequencies. Bayes's Theorem is basically concerned with three probability measures.

Prior probability is the probability of a disease state in the general population. To state it another way, assume that a diagnostician must diagnose a certain organization without receiving any symptom information about the organization. His entire set of knowledge about the situation consists of a table of relative frequencies of each disease state in the general population. The diagnostician's prior probability for any disease state would be the probability of the event's occurrence given that the diagnostician knows nothing about the organization's symptoms. If, for instance, historical data show that five percent of all organizations have disease A, then the prior probability of disease A is .05. Prior probability of disease state i will be designated $P(D_i)$.

The likelihood of a symptom or symptom complex is the probability of the symptom or complex when the organization is known to have a certain disease. Suppose, for instance, that of all organizations that have disease A, it is known that 60% of them have symptom X. Then the likelihood of symptom X given disease A is .6. The likelihood of symptom I

given disease i is written $P(S_j / D_i)$. The likelihood of a symptom complex of n different symptoms is written $P(S_1, S_2, \ldots, S_n / D_i)$.

Posterior probability is the probability of a disease state given the occurrence of a given symptom or symptom complex. Posterior probability is designated $P(D_i / S_1, S_2, \ldots, S_n)$.

Assume that there are m possible desease states, D_i. There are also n possible different symptoms S_j. Then Bayes's Theorem states that the prior probability, likelihood, and posterior probability for any disease state and symptom complex are related as follows:

$$P(D_i / S_1, S_2, \ldots, S_n) = \frac{P(S_1, S_2, \ldots, S_n / D_i)P(D_i)}{\sum_{i=1}^{n} P(S_1, S_2, \ldots, S_n / D_i)P(D_i)}$$

Historical data have traditionally provided the relative frequencies that were used as prior probabilities and likelihoods. Researchers would compile a data bank of diagnosed cases, and from this set of data would develop tables showing the relative frequency of occurrence of each disease state and of each symptom complex for any given disease. With this information, the diagnostician could apply Bayes's Theorem when confronted with a certain symptom complex, arriving at a posterior probability for each possible disease.

Problems With Historical Data and Relative Frequencies

Several problems result from the exclusive use of historical data and relative frequencies in developing the prior probabilities and likelihoods. One of these is that there may be no objective criterion to determine whether or not the diagnoses were correct. The result is a "Catch-22" situation. If objective, reliable diagnosis criteria already exist, then there is no need to develop a new technique for diagnosis. But if such criteria do not exist, then there is no way to judge whether or not the new technique is valid. Ultimately, the value of diagnosis is in its resultant treatment prescription. The objective of diagnosis should be to classify cases (organizations or patients) into different treatment categories, i.e., into classes in which each member of the class reacts to any treatment the same as all other members of the class. Since the historical data may contain incorrectly diagnosed cases, the relative frequencies based on those diagnoses will obviously be incorrect. This makes both the prior probabilities and the likelihoods suspect.

A second problem, perhaps even more severe, is that most data banks are too limited to provide an adequate set of relative frequencies. Unless the sets of diseases and symptoms are severely limited, a very large set of cases is needed to develop accurate relative frequencies.

For instance, consider the limited situation in which there are only four different symptoms under consideration and each symptom can take on only two values. There are 2^4, or 16 different possible symptom combinations. Now assume that there are three possible disease states. This means that there are 16 x 3, or 48 different likelihoods to compute,

one for each symptom complex for each disease. An adequate data base would require several hundred cases in order to be assured that a reasonable number of the cells in the relative frequency table for each disease were non-zero.

Cells whose relative frequency is zero present a special problem. Particularly when the data base is small, there will be a significant number of empty cells. Even the most pure relative frequentist statistician will be quick to admit that not every empty cell should lead the researcher to assume that the probability of that symptom complex is zero. The researcher has no sure way of knowing whether the empty cells should in fact be cells with a zero probability of containing cases, or whether they are cells with a low but non-zero probability, or whether sampling error has caused them to be empty.

This zero-frequency problem will be particularly acute in the current study. The symptoms for the study will be scores on 14 different indexes on the Survey of Organizations. Each index can take on virtually an infinite number of values between one and five. However, in order to develop relative frequency tables, the data will be grouped into intervals. There are at least three ways of defining such intervals.

First, one could simply take the possible range of scores and divide this range into intervals of equal length. This would provide equal raw-score intervals.

A second method would be to obtain standard scores for each case on each variable, and divide the scores by equal standard-score units.

The third way would be to define percentiles, and divide the scores into equal percentile intervals.

Regardless of which interval classification method is chosen, the scores will be grouped into a fairly large number of intervals. For instance, each variable could be grouped into intervals that define the five-percent points of a percentile scale, thereby reducing the set of possible values to 20. This set of 14 indexes, each with 20 possible values, yields a set of 20^{14}, or 6.55×10^{20} possible symptom combinations. Any data bank is going to be too small to provide for more than a very small percentage of non-zero cells in such a symptom matrix.

There is one condition under which the number of cells becomes more manageable. When the symptoms are all conditionally independent, the researcher need not fill in all 20^{16} possible cells. One needs to know only the marginal values, i.e., the relative frequency of each value for each single symptom. In this case there would be 16 x 20, or 320 cells. Furthermore, such a 320-cell matrix would need to be completed for each of the disease states under consideration. In the current study using the results of the H-GROUP classification program there are 17 different profiles or disease states. In order to have even the potential of having every cell be non-zero, one would need 20 cases for each disease state, or 340 cases in the data bank. The actual data bank is considerably larger than 340 cases. This large sample size will provide fairly accurate descriptive statistics on each index (variable) for each disease state, but will still contain a large number of zero-frequency cells.

It is known that the data are not conditionally independent. This can be both an advantage and a disadvantage. The disadvantage lies in the fact that unless one knows precisely the exact relationship among the symptoms, incorporating the knowledge of the non-independence into Bayes's Theorem is impossible. One is left with the problem of estimating

20^{16} cell frequencies. The advantage of knowing that the data are not conditionally independent is that it may enable the observer to reduce the number of symptoms that must be considered in making the diagnosis. For instance, if knowing X provides a great deal of confidence about the value of Y, one need not observe Y, since it will add little new information to the situation.

Suggested Solutions to the Problems

The problem of massive relative frequency tables is faced in virtually every non-trivial diagnosis situation. Such large tables make it almost impossible to create an adequate data base for generating cell frequencies, and are also difficult to apply, since there is such an overwhelming number of cell values to use in applying Bayes's Theorem. Gustafson et al. (1969) suggested that the proper way to handle this problem is through man-machine systems in which man estimates the likelihoods based on his personal knowledge and the historical data base.

A similar solution may be in order in the current situation. In order to use such an approach, however, one must accept the notion of subjective probability that was discussed earlier. Recognizing that the data base that is used to generate the likelihoods is not going to be sufficiently large to provide accurate estimates of cell frequencies, one can use the data base as a random sample from which one can make inferences about the theoretical distribution of cell frequencies.

In other words, the cell relative frequencies become not the likelihoods themselves but clues from which one develops likelihoods. The researcher is seeking for a way of describing a degree of belief about

the occurrence of the empty cells, and is not restricted to a simple frequency count.

For the time being, ignore the problem of non-independence: assume that the symptoms are conditionally independent from each other. For this situation that means that one must determine the likelihoods for 16 x 20 x 17 = 5,440 different symptom-disease cells. In an earlier study (Bowers & Hausser, 1975) using civilian and Navy work groups, each of the 17 different profiles was represented by at least two percent of the groups in the defining samples. In that study, this meant that there were at least 11 groups in each of the 17 profile categories. Although 11 groups is not enough to assure that all, or even most, of the cells in the relative-frequency table are non-zero, it does provide a large enough sample to obtain a fairly good estimate of the mean and standard deviation of each index for each profile type.

Consider each symptom (index) as a random variable that can take on any of a number of values that are at least ordinal in nature and are probably interval measures. Intuitively, it makes sense to assume that for any disease state, the distribution of values of any given index is an orderly one. The historical data base provides a random sample of values of the index, but is subject to all the sampling errors inherent in any random sample. From this point of view, it becomes unnecessary to maintain the artificially imposed interval groupings of the data scores. The groupings were originally done in order to create cells whose relative frequencies were to be estimated. The actual data values, however, fall along a continuous line, and may be considered as having been generated by an underlying continuous probability density function. Now, instead of

looking at cell frequencies, simply observe the mean and standard deviation of the defining sample. From these sample statistics one can infer the characteristics of the underlying data generating population.

The researcher must decide on a family of distribution functions which will be assumed to be the data generating functions for the likelihoods. Two continuous distributions that immediately come to mind are the normal distribution and the beta distribution. Both distributions are well-known, easily described distributions. Whether or not either of them describes adequately the underlying distributions for the indexes will be determined in the course of the research, but using either of them should result in more accurate posterior probability estimates than would the use of inadequate cell relative frequencies.

An example might make the picture clearer. Table 3 illustrates how one might derive likelihoods using the sample mean and standard deviation as a starting point.

Example

Assume that the historical data base has classified 50 work groups into Profile A. For Index X, the distribution of scores is shown in Table 3.

Figure 4 illustrates the frequency distribution for the sample. From the mean and standard deviation of the sample, best-fitting normal and beta distributions can be derived. (See Figure 5 and Figure 6.) From these distributions one then can develop likelihoods for the 20 intervals. Table 4 displays the likelihoods derived from the three methods. Notice that using of either the normal distribution or the beta distribution eliminates the zero-frequency problem, and smooths out the curve.

Table 3

SCORES OF 50 HYPOTHETICAL GROUPS

FORMING SAMPLE FOR EXAMPLE 1

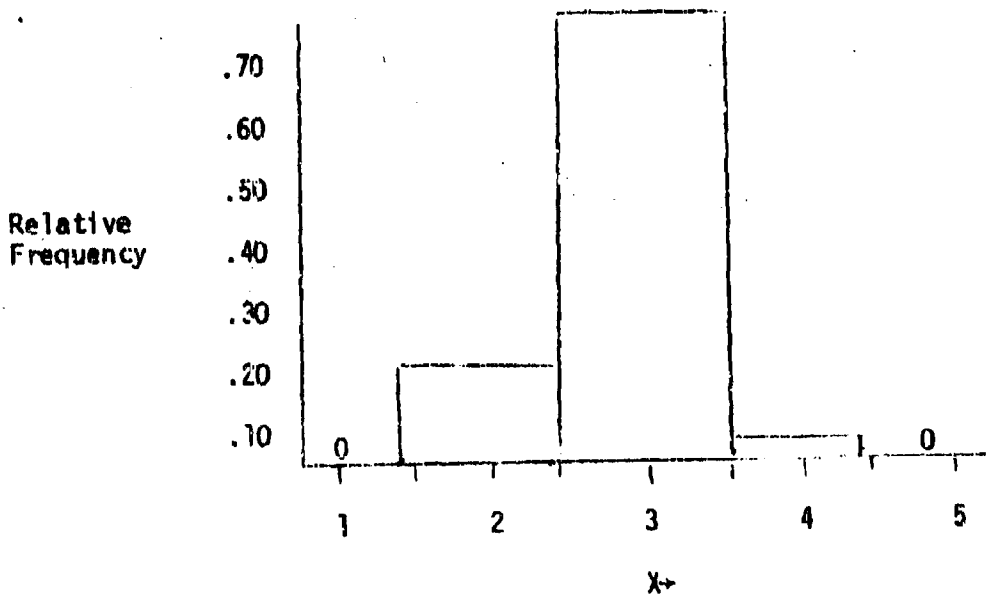
Group Number	Mean Score	Group Number	Mean Score
1	1.71	26	2.81
2	1.95	27	2.83
3	2.30	28	2.85
4	2.32	29	2.86
5	2.36	30	2.87
6	2.37	31	2.91
7	2.40	32	2.94
8	2.44	33	2.94
9	2.48	34	2.99
10	2.49	35	3.03
11	2.51	36	3.10
12	2.51	37	3.12
13	2.52	38	3.15
14	2.53	39	3.16
15	2.56	40	3.18
16	2.62	41	3.24
17	2.63	42	3.25
18	2.64	43	3.26
19	2.65	44	3.27
20	2.65	45	3.31
21	2.67	46	3.40
22	2.67	47	3.54
23	2.69	48	3.60
24	2.70	49	3.71
25	2.77	50	3.92

Mean of Groups = 2.7876

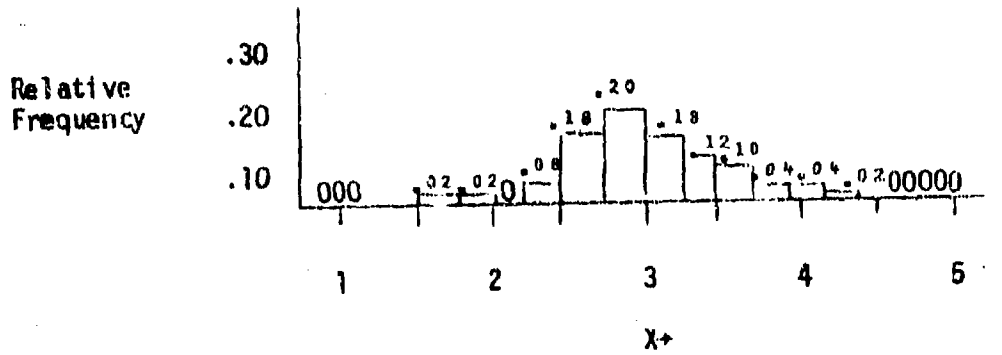
Standard Deviation of Groups = .5527

Figure 4

Histograms of the Relative Frequency of Scores in Example 1



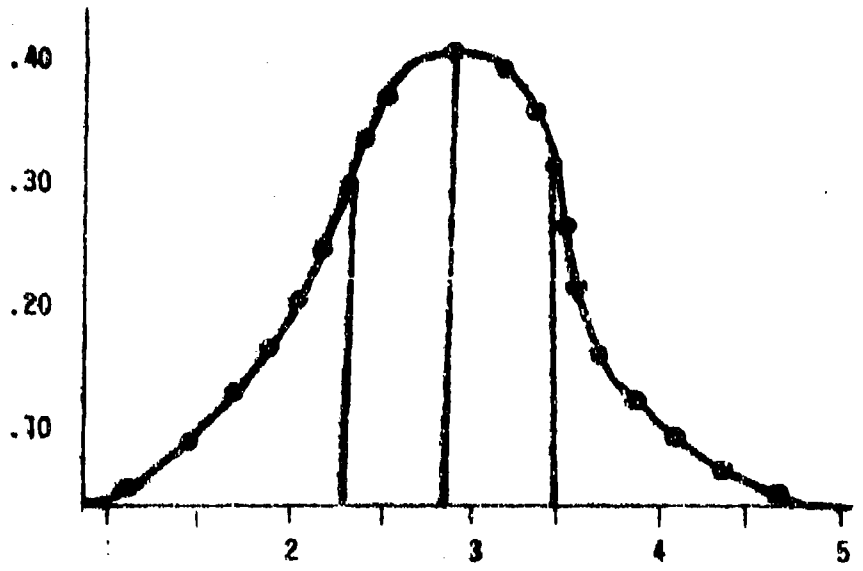
1a. Frequency distribution of five intervals



1b. Frequency distribution of twenty intervals

Figure 5

Normal Distribution Based on The Sample in Example 1



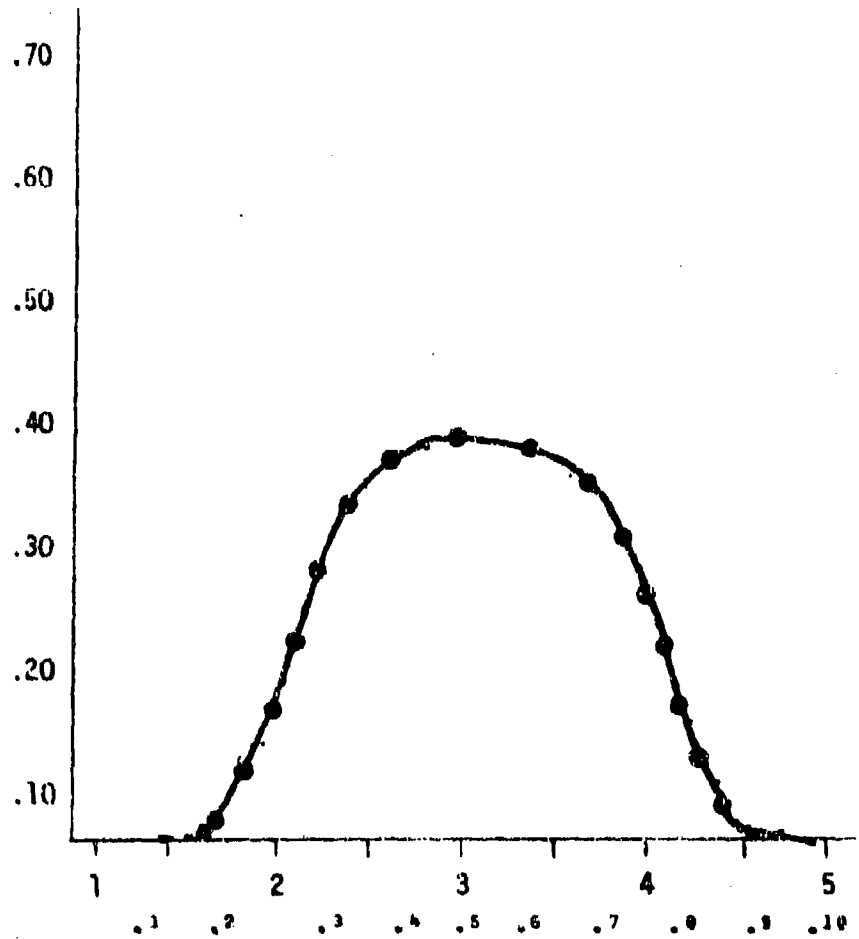
$\bar{X} = 2.78$

$\sigma = .55$

Figure 6

Beta Distribution Based on the

Sample in Example 1



$f(\chi) = K\,\chi^{4.8368}\,(1-\chi)^{5.6051}$

where $\chi = \frac{R-1}{4}$ and R = raw mean score

K is a normalizing constant

Table 4

LIKELIHOODS FOR SCORES IN 20 INTERVALS BASED ON

RELATIVE FREQUENCY, NORMAL DISTRIBUTION , AND BETA DISTRIBUTION

		Method of Calculation	
Interval	Frequency	Normal	Beta
1. 1.0-1.2	0	.0014	.0001
2. 1.2-1.4	0	.0040	.0017
3. 1.4-1.6	0	.0098	.0082
4. 1.6-1.8	.02	.021	.023
5. 1.8-2.0	.02	.040	.047
6. 2.0-2.2	0	.068	.075
7. 2.2-2.4	.08	.097	.103
8. 2.4-2.6	.18	.125	.124
9. 2.6-2.8	.20	.141	.134
10. 2.8-3.0	.18	.140	.131
11. 3.0-3.2	.12	.125	.115
12. 3.2-3.4	.10	.093	.092
13. 3.4-3.6	.04	.063	.066
14. 3.6-3.8	.04	.037	.042
15. 3.8-4.0	.02	.019	.023
16. 4.0-4.2	0	.0091	.010
17. 4.2-4.4	0	.0035	.0036
18. 4.4-4.6	0	.0012	.0008
19. 4.6-4.8	0	.0004	.0001
20. 4.8-5.0	0	<.0001	.00003

Dealing With the Problem of Non-Independence

The problem of non-independence of symptoms is a difficult one to deal with in a Bayesian framework. Gustafson et al. (1969) suggested that in the medical situation the problem could be alleviated by having the diagnosing physicians cluster the data into groups that they perceive as being relatively independent. Again, a somewhat similar procedure could be used in the present study. It is straightforward to obtain a matrix of pairwise correlations among the 14 indexes used in the organizational diagnosis. From such a correlation matrix, the indexes can be grouped into relatively independent clusters. When data are conditionally independent, the application of Bayes's Theorem is straightforward.

Let $P(S_j / D_i)$ be the likelihood for a given symptom cluster, and assume that each symptom cluster is independent of all others. Let $Q(S_j)$ be the sum over all disease states (indexes) of the $P(S_j/D_i)P(D_i)$. In other words, $Q(S_j) = \sum_{i=1}^{n} P(S_j / D_i)P(D_i)$. Then the posterior probability of disease state i when only symptom complex j has been observed is:

$$P(D_i / S_j) = \frac{P(S_j / D_i)P(D_i)}{Q(S_j)}$$

Now if one observes not just symptom complex j but all m independent symptom complexes, the posterior probability becomes:

$$P(D_i / S_1, S_2, \ldots, S_m) = \frac{P(S_1 / D_i)P(S_2 / D_i)\ldots P(S_m / D_i)P(D_i)}{Q(S_1)Q(S_2)\ldots Q(S_m)}$$

Creating such conditionally independent clusters thus simplifies the computation of posterior probabilities once the likelihoods of the clusters have been determined.

The diagnostician must, however, still deal with the issue of non-independance within each cluster. One way to deal with it would be to compute a single score for the cluster, an average of the scores for the indexes comprising the cluster. This is simply a data reduction method.

A second way of dealing with this issue would be to look at the multiple correlations of related items, and reduce the likelihoods based on those correlations. For instance, suppose that Index A and Index B have a correlation coefficient of .70. This means that .49 of B's variance is accounted for by variability in A. If the two indexes were conditionally independent, one would obtain the likelihood of the AB combination by simply multiplying the likelihoods for the two indexes. But since the two are correlated, multiplying the two likelihoods will yield a likelihood for the cluster that is too large. That over-estimation can be compensated for by reducing the individual likelihoods by the amount of variability that they share. Thus, if the likelihood ratio for two profiles was X if the symptoms were independent, then that likelihood ratio would be .51X if they were correlated by .70. This is because the common variance = r^2, or .49. Unique variance of each is 1.00 - .49, or .51.

A word about likelihoods and likelihood ratios is in order. One of the more difficult values to obtain in using Bayes's Theorem is the denominator of the equation, $\sum_{i=1}^{n} P(S_1, S_2, \ldots, S_n / D_i)P(D_i)$. When Bayes's Theorem is written in what is called the odds-likelihood ratio form, that denominator drops out. Odds are simply ratios of probabilities.

The prior odds of disease i against disease k are simply the ratio of the prior probabilities of the two diseases. Thus, $\Omega_0 = \frac{P(D_i)}{P(D_k)}$. Similarly, likelihood ratio for two disease states is simply the ratio of the likelihoods for the two diseases. And posterior odds are the ratio of posterior probabilities. Using the odds-likelihood ratio form of Bayes's Theorem simplifies things a great deal. Let Ω_0 be the prior odds of disease i against disease k. Let Ω_1 be the posterior odds, and L_j be the likelihood ratio for symptom complex j. Then Bayes's Theorem becomes: $\Omega_1 = L_j\Omega_0$. When one is dealing with conditionally independent symptom complexes, the equation becomes: $\Omega_1 = L_1L_2 \ldots L_n\Omega_0$.

In the preceding paragraph it was suggested that the likelihood ratio of a symptom in a given symptom cluster be reduced by the amount of overlap that it had with others in the cluster. Applying this principle to likelihoods rather than likelihood ratios would be difficult because of the complexity of computation, but is straightforward when one uses the odds-likelihood ratio form of Bayes's Theorem.

Specific Proposals for the Bayesian Diagnosis

The specific analyses under consideration suffer from each of the difficulties that can occur in trying to apply a Bayesian approach to organizational diagnosis. However, the techniques described previously can be used to alleviate the difficulties. It is proposed that four different tests of the Bayesian technique be made, two of them assuming conditional independence and two of them dealing with issues of non-independence. In all cases, however, it is proposed that the data bank be divided into a defining sample and a validation sample. The defining sample should consist of approximately half of the work groups in the

data bank. Thus the defining sample N will be approximately 3,000. Classify the 3,000 work groups using the criteria developed by the H-GROUP technique. Create for each profile type a set of descriptive statistics and relative frequency tables for each of the 16 indexes.

Tests Assuming Independence

For this part of the research, assume that the 16 indexes are conditionally independent. Based on that assumption, the researcher is concerned only with determining the probability density function of each index for each profile, and need not consider the problem of determining combined density functions for all 16 indexes. Further, it is proposed that the researcher use the odds-likelihood ratio form of Bayes's Theorem, which means that the ratios of the density functions rather than the absolute values of the density functions are the values of interest. The two different ways of developing likelihood ratio density functions will involve two different families of continuous distributions.

Method 1: The Family of Normal Distributions. For each profile, determine the mean and standard deviation of the member work groups on each index. Assume that the underlying data generating distribution is a normal distribution with mean and variance equal to the sample mean and unbiased estimate of the sample variance. The result will be for each index a set of normal density functions, one for each of the 17 profiles.

In order to use the odds-likelihood ratio form of Bayes's Theorem, select one profile of the 17 as the "standard," against which the remaining 16 profiles will each be compared. Then to obtain posterior

probabilities across the 17 profiles for any work group that one wants to diagnose, simply apply the odds-likelihood ratio form of the theorem 16 times, and convert the posterior odds to probabilities. Under the assumption of independence, the posterior odds for any profile comparison will simply be the product of all the likelihood ratios (one likelihood ratio for each index) times the prior odds.

Classify each of the defining sample work groups using the prior odds and likelihood ratios developed using the normal density function assumption. A further test would be to classify the remaining work groups, those in the validation sample.

Method 2: The Family of Beta Distributions. The beta distribution is a unimodal, continuous distribution, as is the normal distribution. It has some characteristics that are particularly appealing in the current situation. The range of scores on the Survey of Organizations is limited; values may range between one and five. The normal distribution assumes an infinite range, so it is known a priori that the normal distribution can never exactly fit the actual distribution of work group scores. This discrepancy becomes especially severe when the mean of the distribution is close to either end of the range, since the actual distribution must be skewed but the normal distribution is symmetric.

The beta distribution, however, is defined over a closed interval, such as that which the 500 scores cover. It reflects the skewness that is required by such a range restriction. The beta distribution is defined by two independent parameters which are uniquely determined by the mean and variance of the distribution. In straightforward fashion one may compute

the two parameter of the beta distribution, thus obtaining for each index on each profile the underlying density that will determine the likelihood ratios.

Once the density functions have been determined, proceed as for the normal distribution test, classifying each work group in the defining sample and the validation sample according to the posterior odds resulting from the beta-family likelihood ratios.

Tests Assuming Non-Independence

Two methods will be used to test the Bayesian diagnosis technique which take into consideration the issue of conditional non-independence. These methods should be applied after the two methods assuming independence have been done. Select the better family of distributions to generate the underlying probability density functions.

For both methods, the first step is to obtain a correlation matrix of all 16 indexes. From this matrix, select clusters that are relatively uncorrelated with each other. This procedure should result in the reduction of 16 non-independent indexes to about five relatively independent clusters.

Method 3: Creation of Grouped Scores. This method deals with the non-independence problem by reducing the data to a set of hopefully independent scores. After the clusters have been determined, create a single score for each cluster that is the mean of the scores of the indexes belonging to the cluster. Then treat the new scores as independent data and develop the distribution functions as described in the section assuming independence. Once the distribution functions have been determined, carry out the classification tests as described previously.

Method 4: Reduction of Likelihood Ratios by Degree of Shared Variability. Starting with the relatively small number of independent index clusters, look at the multiple correlations among them. This can be done by selecting as the "dependent" variable the index which has the highest correlation to all the others. Then, treating the other indexes in the cluster as independent variables, perform a multiple regression. One of the outputs of the multiple regression is the partial correlation of each variable from which one can compute the amount of variance uniquely accounted for by each variable. The likelihood ratio for a given cluster would be the product of each index likelihood ratio, reduced to the amount of unique variance it contributes to the overall score.

Example

Suppose that Cluster A is composed of Index X, Index Y, and Index Z. Also suppose that Index X is the most highly correlated with each of the other two indexes. Make Index X the dependent variable. Suppose that the multiple regression of X with Y and Z results in a multiple R of .86 and partial correlations of .69 and .08 for Y and Z respectively. Then assume that the marginal likelihood ratios for X, Y, and Z are L_X, L_Y, and L_Z. Then the combined likelihood ratio for the cluster would be:

$L_C = (.86)^2 L_X \times (69)^2 L_Y \times (.08)^2 L_Z$.

The multiple regression need not be done for each profile, but should be done only once, for the overall defining sample. This will then yield reduction factors for each index cluster that will be used in all likelihood ratios involving that cluster.

This procedure will result in downgraded likelihood ratios for each cluster. These modified likelihood ratios will then be treated as conditionally independent figures in Bayes's Theorem.

Why Use the Bayesian Approach

It is apparent that in any non-trivial situation, the use of Bayes's Theorem to determine the probabilities of the possible profiles given the index values is difficult at best. The two most difficult problems are the almost guaranteed inadequacy of the data base and the issue of conditional non-independence. Modifying the technique so as to in some way alleviate these problems leads to less confidence in the final probabilities, since they are based on subjective estimates that are only "best guesses" about the veridical values. If the measure of goodness of the technique is its distance from the ideal profiles as developed by H-GROUP, then the Bayesian approach will assuredly perform less well than a distance-based approach.

If, however, other criteria of goodness are used, the Bayesian approach may prove to be more useful than some other approaches. One positive aspect of the Bayesian approach is that it does not, strictly speaking, reach a diagnosis at all. The final output is a set of probabilities over all possible profiles. Presumably, the diagnostician would select that profile that has the highest probability. If, however, the ultimate use of the diagnosis is in determining treatment, the diagnostician may want to do more than make a diagnosis. He may want to select the treatment that will yield the highest expected return. In such a case, he should not ignore all the less likely diagnoses, but should weigh the expected values of the various treatments by the likelihood that they are to be applied to each of the possible disease states.

Another criterion of goodness may be the ability of a practitioner to make a diagnosis while in the field. If he is restricted to a distance-type model, that may be difficult if not impossible. With the Bayesian model, determining the initial set of likelihood ratios is the difficult part; once the values have been determined, however, it would be fairly simple to equip the practitioner with a simple Bayesian equation to use in the field. It would proably require the use of a hand calculator, but not a computer.

There are various other criteria, such as computing expense, the amount of data that must be evaluated in order to reach a diagnosis, and the ease of selecting second and third choices. Whether or not the Bayesian approach would perform well under those criteria is certainly testable.

EXAMINATION OF TECHNIQUES

Accuracy Based Criteria

The purpose of this section is to present five different criteria to use in evaluating classification procedures, each of which is based on a measure of accuracy. In later sections, non-accuracy based criteria are discussed.

The four different classification procedures to be investigated are a decision tree (DT), multiple discriminant function (MDF), a Bayes rule (B), and a straight (least squares) distance function (DF). The typology into which work groups are to be assigned has been discussed earlier in this report. Two aspects of the typology and its development are relevant to the investigation here. First, it has been derived by a clustering algorithm which, in effect, groups the observations (vectors of mean scores from work groups) so as to minimize the variance within clusters. The variance metric is also a distance metric, which will result in the same classification by the DF for a particular work group as you would get by including that work group in the clustering process originally. The consequence is that we will use the distance function (DF) classification as the correct classification for any particular work group.

The typology developed by the clustering algorithm allows the creation of a vector of scores from the averages across all work groups within the cluster (Bowers & Hausser, 1975). Thus, we can think of the typology containing 17 types as a set of 17 vectors. That is, a type is represented by a single vector of scores.

Proportion of Correct Classification

The classical criterion for evaluating classification schemes is the proportion of agreement with some external "expert" opinion. In the present situation, as discussed above, the expert opinion will be that provided by the distance function (DF). However, there is some interest in seeing how well one data analytic technique is able to reproduce another data analytic technique when additional criteria (beyond proportion of agreement) are considered. The reader is referred to the following sections for discussion of other criteria.

As a beginning, we propose to analyze the three classification procedures, MDF, B, and DT, by comparing the proportions of agreement with DF. The straightforward process will be as follows:

> For each of MDF, B, and DT, compute the proportion of the (approximately) 3,100 work groups which are assigned to the same type by both the given procedure and DF.

The figures obtained from the above analysis can be viewed, for each procedure, as the percentage correct within each type of the typology as well as aggregated across all types. Such a review may enhance the application of the technique in other settings.

Reproduce the Typology

Assume a particular classification procedure has classified, say, k groups into a single type within the typology. It then would be possible to compute the averages for those k groups, resulting in a single vector of index scores. It is desirable to have this vector of scores for the

assigned groups be close to the vector of scores which represent this type. A measure of accuracy of prediction would be to compute the sum of the distances between each of the 17 types and their corresponding mean vectors from the groups assigned to them.

A criterion for evaluating the different classification schemes, then, would compare the sums of distances between the typology vectors and the group means. Because of the use of a standard distance measure to compare the type vector and the vector of average scores from the observations assigned to that type, the distance function will do best here, as well. However, it would be appropriate to compute the sum of the distances for DF to use as a standard against others which could be compared.

The procedure to implement this process, using all (approximately) work groups, is as follows:

For each of MDF, B, DT, and DF:

1. Compute a vector of average values for each set assigned to a type by the procedure.
2. Compute the distance between this vector of averages and the type vector.
3. Sum the distances across all types.

For each of MDF, B, and DT:

4. Compute the ratio of the sum to the sum for DF.

Weighted Dimensional Distances

The classical distance metric considers each of the dimensions equally. It is known that of the 14 dimensions resulting from the SOO they are not all of equal importance. Merited, then, is the consideration of a distance metric which weights the dimensions according to their importance.

One possible scheme for ordering the 14 dimensions would be to use the criterion of the strength of the relationship between the single dimension and a measure of productivity. We will refer to such a rank-ordering of the dimensions, and the weights assigned to such an ordering, as the performance ordering and performance weights, respectively. Another example of a method for ordering the dimensions is to do so according to the susceptibility to change. The different dimensions of organizational functioning as defined by the SOO respond differentially to planned change activities, and some are more difficult to change than others. The weights resulting from ordering according to this criteria will be called change weights.

Algebraically, the classical distance is given by $d = \sqrt{\Sigma(y_i - Z_i)^2}$ where y and Z are the two points of interest. The weighted distance is given by $d_w = \sqrt{\Sigma w_i (y_i - Z_i)^2}$, where $w_1, w_2, \ldots, w_n$ are the weights assigned to the dimensions.

The actual choice of the weights can be based on several possible criteria. The underlying premise is that the weights reflect the relative importance of the dimensions. As mentioned above, it is possible to rank order the 14 dimensions of the SOO according to their correlation, pairwise, with some productivity criterion. One such criterion is total variable

expense (TVE), measured as a percentage, which reflects actual costs compared to a standard. Data are available on the correlations between work group SOO scores and TVE, for a large number of work groups in a variety of organizations. These correlations, r_i (between the ith index & TVE) can be used as weights according to $w_i = \frac{r_i}{c}$, where c is a constant chosen so that $\Sigma w_i = 1$.

Considerable research has been done on changing organizations, with specific measurement of that change by the SOO (Bowers & Hausser, 1977; Bowers, 1973; Franklin, 1976). As a consequence of the work done in these past investigations, we are able to obtain change scores on each of the 14 SOO indexes over a large number of work groups without much additional work. Let these change scores be represented by $g_1, g_2, \ldots, g_{14}$. Define $w_i = g_i/c$, again where c is chosen so that $\sum_{i=1}^{14} w_i = 1$. These then will be the change weights.

The procedure to compare the classification techniques will be as follows:

1. Establish a set of performance weights and a set of change weights.
2. Compute vectors of average values for each set assigned to each type by each classification procedure.
3. Compute the weighted distance, d_w, for each set of weights, between the average for the type and the type vector.
4. Sum the weighted distance across the 17 types, for each procedure, and for both sets of weights.

It is noted that this procedure is similar to that suggested under the heading, Reproduce the Typology.

Zero-One Count

In assigning a work group to a single type within the typology, two different patterns of work group scores may lead to the same distance from the type. Consider the following example, simplified to reflect only four dimensions.

type	(2,2,2,1)	
work group 1	(4,4,4,3)	distance = 4
work group 2	(2,2,2,5)	distance = 4

One approach to overcoming the above situation is to use the weighted dimensional distances described earlier. An alternative approach is to count the number of dimensions for which the assigned group is significantly different from the type. If we define significantly different as being greater than or equal to two in the above example, the counts would be four for work group 1 and one for the second work group. Mathematically, this counting can be represented as the sum of the values c_i,

$$c_i = \begin{cases} 1, & \text{if } |Y_i - X_i| \geq s \\ 0, & \text{if } |Y_i - X_i| < s. \end{cases}$$

Here s is the value defining significantly different, and $\underline{Y}$ and $\underline{X}$ are the vectors representing the points of interest. We will let $C = \Sigma c_i$. Then C may be computed for each work group assigned to a particular type, and either an average computed for each type, or summed across all work groups and types, called CT. The choice of s should be based on some measure of relative magnitude or theoretical argument. For the present study, we propose s be chosen as a measure of the average variability of the SOO indexes. Letting $\bar{q}_i^2$ be the variance within types of the ith index averaged across types. Then let $s = a \cdot \sqrt{\frac{1}{14} \sum_{i=1}^{14} \bar{q}_i^2}$, where a may take on the values 1, 1.5, and 2.0.

Thus, CT could serve as a criterion for evaluating the accuracy of classification. The procedure involved would be:

1. Assign a value to s, the significant difference.
2. For each procedure, compute C for each work group.
3. Compute CT for each procedure.

Alternatives could include examining different values of s, or even different values for each dimension, say s_i. This latter possibility has the potential for weighting the dimensions.

Severity of Misclassification

One argument against the use of the classical frequency of correct classifications is that it treats all misclassifications as equal. That is, there is no distinction between misclassifying a particular work group into any of the k-1 incorrect types in a typology of k types. It is not unusual, however, for the cost of misclassification to be widely different across the k-1 incorrect types, as well as being contingent on the correct type. For example, incorrectly diagnosing an individual with a severely sprained ankle as having a broken ankle or having leukemia has different costs; additionally, incorrectly diagnosing an encephalitis case as a broken ankle or leukemia has yet different costs.

An alternative to the straight proportion of correct classification criterion of accuracy is one which allows for differential costs of misclassification. In particular, we propose two different costing models for misclassification for inclusion in the present study.

1. That a cost of 0 be assigned to any misclassification of a work group into a category which calls for (a) treatment if so does the correct category, or (b) non-treatment if so does the correct category; otherwise the cost is 1.
2. That a cost of 0 be assigned to any misclassification of work group into a category which calls for a treatment which is known to have a positive effect on the correct category, and a cost of 1 if the incorrectly selected category calls for a treatment which is known to have no effect, or a negative effect, on the correct category.

These are only two models of a vast number of possibilities for costing misclassification. However, they are reflective of the primary concerns of misclassification -- the application of inappropriate or harmful treatments.

Bowers and Hausser (1977) examined the effects of different change strategies on each of the 17 groups of the typology. They rated the effects either as negative, neutral, or positive. For number 1 above, it is possible to define a work group calling for treatment. Otherwise, the work group calls for no treatment. The results of Bowers and Hausser (1977) can also be used to establish the appropriate pattern of costs for number 2 above.

Notationally, the above procedures can be generalized as follows: let the typology contain k types. Let C_{ij} represent the cost of diagnosing type i into category j. Then $C_{ii}=0$, but we need not have C_{ij} equal. The cost matrix is then a k x k matrix, given by:

	column represents classified state				
	C_{11}	C_{12}	C_{13}	. . . .	C_{1k}
row represents true state	C_{21}	C_{22}	C_{23}	. . . .	C_{2k}
	.	.	.		.
	.	.	.		.
	.	.	.		.
	C_{k1}	C_{k2}	C_{k3}	. . . .	C_{kk}

The traditional accuracy criterion of frequency of correct classification has cost matrix

$$\begin{matrix} 0 & 1 & 1 & \ldots. & 1 \\ 1 & 0 & 1 & \ldots. & 1 \\ 1 & 1 & 0 & \ldots. & 1 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 1 & 1 & 1 & \ldots. & 0 \end{matrix}$$

For procedure 1 above, assume that the first t types call for treatment, and that the last k-t types indicate no treatment. (We may consider treatment as a major intervention into the life of the work group.) Then the cost matrix for procedure 1 is given by

$$\begin{array}{c} \\ t \\ \\ \\ \\ \\ k-t \\ \\ \\ \end{array} \begin{array}{c} \overbrace{}^{t} \quad \overbrace{}^{k-t} \\ \left[\begin{array}{cccccc} 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ . & & & & . & \\ . & & & & . & \\ . & & & & . & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ 1 & 1 & 1 & \ldots & 1 & 0 \end{array}\right. \end{array} \begin{array}{c} 1\ldots1 \\ 1\ldots1 \\ \\ \\ \\ 1\ldots1 \\ 0\ldots0 \\ 0\ldots0 \\ \\ \\ \\ 0 \end{array}$$

If we let f_{ij} be the frequency (relative) of classifying true type i as type j, the accuracy index A is given by $A = \sum_{i,j} C_{ij} f_{ij}$ which one wishes to minimize.

For the classical procedure, A is just the total frequency of misclassification.

A second variation in the cost of misclassification is naturally available using a Bayes classification procedure. When using such a procedure, the result is a set of k (posterior) probabilities that the work group is of type i, i = 1, 2, ...,k. The generally accepted process is to classify the work group into the type for which the probability is the largest. However, it is possible to say that, with much higher probability, tha he work group belongs to one of the three, say, types with the three largest probabilities. (Clearly the selection of three is arbitrary.)

Notationally, the above can be developed sequentially from the earlier model. Let b_{nij} be the cost of classifying the n^{th} work group, which is actually of type i, into type j. (Note that we don't have b_{ni_2} and b_{ni_1}, but only b_{nij} for one pair of values i, j.) In effect, $b_{nij}=C_{ij}$. Then we have

$$\sum_{n=1}^{\#\text{ work groups}} b_{nij} = \sum_{i,j} C_{ij} f_{ij}.$$

If one wants to consider the two most likely groups of classification, then, notationally, let

$$b_{nijk} = \begin{cases} \dfrac{C_{ij} + C_{ik}}{2}, & j \neq i \\ & k \neq i \\ 0, & \text{otherwise.} \end{cases}$$

and the three most likely groups procedure has the following notation:

$$b_{nijkl} = \frac{C_{ij} + C_{ik} + C_{il}}{3}, \quad \begin{matrix} j \neq i \\ k \neq i \\ l \neq i \end{matrix}$$

$$= 0 \qquad \text{otherwise}$$

Here b_{nijkl} represents the cost for actually having classified the nth work group, which is of type i, as type j or k or l. The cost is zero if any of j, k, or l are correct. Clearly one is interested in minimizing

$$B = \sum_{n=1}^{\#\text{ work groups}} b_{nijkl}$$

While the Bayes approach lends itself naturally to classifying a work group as being one of a set of three types (rather than a single type), it is possible to do this with other classification procedures as well. For example, with the distance statistic, which assigns a work group to the type to which it is closest, it is possible to define a set of three types by choosing the three types to which it is closest.

In that the approach by definition gives the right answer, it is appropriate to ask why classify into a set of three types. However, as discussed elsewhere in this report, other factors such as reduced data sets, ease of computation, etc. lead to continued consideration of using the distance procedure in this way.

We then define a third cost of misclassification procedure as

3. That a cost of 0 be a signed to any classification of a work group into a set of three types where one of the three is the correct type, and otherwise, the cost is 1.

Non-Accuracy Based Criteria

Information Required to Make a Decision

One criterion by which one can compare diagnostic techniques is the information required to reach a diagnosis. If two techniques perform equally well when the full set of indexes is used in the diagnosis, then it will be to the diagnostician's advantage to use the one which requires less information to reach a decision. Collecting and processing information is costly, in terms of money, time, and complexity of processing. For instance, if one can obtain results using three pieces of information that are as good as the results using five pieces of information, it will clearly be advantageous to use only three pieces of information.

The most obvious way of reducing the amount of required information is to find an appropriate way to reduce the number of indexes that are used in the diagnostic process. There are two ways of reducing the number of indexes. The first is to discard or eliminate indexes that have been shown to be unnecessary for the diagnosis. The second is to combine several indexes into "super-indexes." For instance, it might be possible to combine the four peer leadership indexes into a single index.

First, consider reducing information by deleting selected indexes. In order to test the feasibility of reducing the information required to reach a diagnosis in that way, the following steps are proposed, using the defining sample from the data set:

1. For the Multiple Discriminant Function, set a criterion cut-off point for the weights of the variables. Eliminate from the diagnostic equation those indexes whose weights are below the cut-off point. Classify the defining sample groups based on the abbreviated equation. Compare the H-GROUP results and the complete equation MDF results, using proportion of correct classification as the criterion. If time and resources allow, this could be a step-wise procedure, in which the indexes are deleted one at a time, until a proportion-of-correct-classification cut-off has been passed.

2. For the Bayesian method, the indexes to be eliminated would be those that discriminate least among profile types, i.e., those whose likelihoods do not vary greatly among profile types. Each index has associated with it a distribution of likelihoods for every profile type. The current study proposes to determine those likelihoods not from the relative frequencies of cases in the cells, but by deriving the assumed underlying distributions characterized by the mean and standard deviation of the work group scores in each profile type. This means, for instance, that Index A has associated with it a mean and standard deviation of group scores for each of profile one, profile two, profile three, etc. One measure of the degree of dissimilarity among the profiles is the standard error of the mean, or in other words, the standard deviation of the mean scores across profiles.

The rule for elimination of indexes would be to select a criterion score, and to discard all indexes whose standard error falls below that cut-off.

As was suggested for MDF, the criterion would be the proportion of correct classifications using the degraded equation. The procedure could also be applied step-wise, if resources should permit.

3. For the Decision Tree technique, the selection of easily deletable indexes should be apparent as the tree is constructed. Again, one would select those indexes which least accurately discriminate among the profile types. The degraded decision tree would be tested using the proportion-of-correct classification criterion. This will be discussed more fully in the technical report on the decision tree.

The second method of index reduction is the combining of indexes. In order to test the feasibility of combining indexes, one would need to go back to the data set and perform some clustering procedures. With new clusters, one would again perform H-GROUP to derive a new set of classification equations. Since this would be a very large study in itself, we plan not to look at information reduction of this type in this project.

Amount of Data Required to Develop the Diagnostic Process

Of interest to the researcher is not only the amount of data that must be processed to obtain a diagnosis, but also the amount of data required to generate accurate diagnostic processes.

All of the techniques to be tested require some kind of historical data base, from which the diagnostic process is generated. In all instances, larger data bases should provide more accuracy than smaller ones. If one technique can be generated quite accurately from a smaller data base than another, that technique would be preferable, since it would be more cost effective.

In order to test the relative strength of each technique on this criterion, we propose to draw a random sample (25%) from the defining sample, developing the MDF, Bayes, and Decision Tree models from it. We will then compare the models against the broader based models to test relative efficiency.

Ease of Calculation

The ideal diagnosis technique is not only accurate, but relatively simple to perform. In the organizational diagnosis situation, the diagnostician may very well be a change agent who must make treatment recommendations while on site or otherwi out of contact with computer facilities.

Three ease-of-calculation factors are:

1. Can be calculated on-site versus in a central location.
2. Can be done with a hand calculator or by hand versus requiring EDP facilities.
3. Few things to be calculated versus many things to be calculated.

While not identical, these three factors are obviously not completely independent of one another.

The comparison of diagnostic techniques on these factors will be a matter of subjective judgment. It is proposed that after the optimal solutions have been determined for each diagnosis technique, the researchers create a table indicating their assessment of each technique on the ease of calculation factors. The diagnostician will then be able to select among the techniques when ease of calculation is an important aspect of a project.

SUMMARY

This report both introduces the topic of research on organizational diagnosis and discusses the methodological issues involved in it. It then describes the analyses to be reported in subsequent technical reports in the series.

Any objective review of the literature of organizational development and change reveals that organizational diagnosis is less than a consummate skill in professional practice. Findings from the general field of assessment and classification indicate that an acceptable procedure for organizational diagnosis ought be statistical, rather than clinical or judgmental.

Diagnoses are really probability statements comparing a sequence of observed symptoms or characteristics with hypothesized symptom sequences derived from past research.

The Survey of Organizations data bank provides a ready resource for systematically examining different statistical methods of organizational diagnosis. To its approximately 5,600 civilian work groups, representing a broad array of industrial and governmental settings and levels, are added 520 military work groups drawn in past studies from the Navy and Army.

Four methods of diagnostic classification are to be examined: distance function, multiple discriminant function, Bayesian, and decision tree.

Succeeding technical reports in the series cover the following analyses:

1. Assignment of all 6,119 groups to one of the 17 types in the Bowers and Hausser typology by a distance function method. This procedure will constitute the "expert diagnosis" criterion.
2. Calculation of discriminant function weights from a developmental random half-sample of 3,000 (approximately) groups.
3. Development of likelihood functions to use in Bayes's Theorem, under the following conditions:
   (a) Independence of index measures:
      (i) Assuming an underlying normal distribution.
      (ii) Assuming an underlying beta distribution.
   (b) Non-independence of index measures.
4. Assignment by a decision tree procedure.
5. Comparison of the techniques in terms of accuracy-based criteria:
   (a) Proportion of correct classification for MDF, Bayes, and DT methods compared to a distance function criterion.
   (b) Reproduction of the typology.
   (c) Weighted dimensional distances.
   (d) Zero-one count.

6. Comparison of the techniques in terms of severity of misclassification, defined as costs associated with inappropriate treatment recommendations.
7. Comparison of the techniques in terms of information required by (a) eliminating indexes, (b) collapsing indexes, and (c) sampling of respondents.
8. Comparison of the techniques in terms of ease of calculation.

References

Birnbaum, A. & Maxwell, A.E. Classification procedures based on Bayes' formula. Applied Statistics, 1960, 9, 152-169.

Bowers, D.G. OD techniques and their results in 23 organizations: The Michigan ICL Study. The Journal of Applied Behavioral Science, 1973, 9 (1).

Bowers, D.G. Organizational Development: Promises, performances, possibilities. Organizational Dynamics, 1976a, 4 (4),51-62.

Bowers, D.G. Systems of organization. Ann Arbor, Michigan: University of Michigan Press, 1976b.

Bowers, D.G. Organizational diagnosis: A review and a proposed method. Technical Report to Office of Naval Research, 1974.

Bowers, D.G. & Hausser, D. Group types and intervention effects in organizational development. Technical Report to Office of Naval Research, 1975.

Bowers, D.G. & Franklin, J.L. Survey-guided development: Data based organizational change. LaJolla, Calif.: University Associates, 1977.

Bowers, D.G. & Hausser, D.L. Work group types and intervention effects in organizational development. Administrative Science Quarterly, 1977, 23, 76-94.

Bowers, D.G. & Spencer, G.J. Structure and process in a social systems framework. Organization and Administrative Sciences, 1977, 8 (1), 13-21.

Burke, W. Organization development. Professional Psychology, 1973, 4, 194-200.

Cronbach, L.J. Essentials of psychological testing. New York: Harper & Row, Publishers, 1960.

Crooks, J., Murray, I.P.C., & Wayne, E.J. Statistical methods applied to the clinical diagnosis of Thyrotoxocosis. Quarterly Journal of Medicine, 1959, 28, 211-234.

Fleiss, J.L., Spitzer, R.L., Cohen, J., & Endicott, J. Three computer diagnosis methods compared. Archives of General Psychiatry, 1972, 27, 643-649.

Fox, D.J. & Guire, K.E. Documentation for MIDAS. University of Michigan, Statistical Research Laboratory, 1976, 3rd ed.

Franklin, J.L. Characteristics of successful and unsuccessful organization development. Journal of Applied Behavioral Science, 1976, 11, (4), 471-492.

Gustafson, D.H., Edwards, W., Phillips, L.D., & Slack, W.V. Subjective probabilities in medical diagnosis. I.E.E.E. Transactions on Man-Machine Systems, September, 1969, 10, (3), 61-65.

Holland, J.H. Adaptation in natural and artificial systems. Ann Arbor, Michigan: The University of Michigan Press, 1975.

Kahn, R.L. Organizational development: Some problems and proposals. Journal of Applied Behavioral Science, 1974, 10, 485-502.

Lawrence, P.R. & Lorsch, J.W. Developing organizations: diagnosis and action. Reading, Mass.: Addison-Wesley Publishing Co., 1969.

Ledley, R. & Lusted, L. Reasoning foundations of medical diagnosis. Science, 1959, 130, July 14, 9-21.

Levinson, H. The clinical psycholgist as organizational diagnostician. Professional Psychologist, 1972, 3, 34-40.

Levinson, H. Response to Sashkin and Burke. Professional Psychologist, 1973, 4, 200-204.

Likert, R.L. New patterns of management. New York: McGraw-Hill, 1961.

Likert, R.L. The human organization. New York: McGraw-Hill, 1967.

Likert, R.L. & Likert, J.G. New ways of managing conflict. New York: McGraw-Hill, 1976.

Likert, R.L. Past and future perspectives on System 4. Academy of Management remarks, 1977.

Meehl, P.E. Clinical versus statistical prediction. Minneapolis, Minn.: University of Minnesota Press, 1954.

Melrose, J.P., Stroebel, C.T., & Glueck, B.C. Diagnosis of psychopathology using stepwise multiple discriminant analysis. Comparative Psychiatry, 1970, 11, 43-50.

Michaelsen, L. A methodology for the studies of the impact of organizational values, preferences, and practices on the All-Volunteer Navy. Technical Report to Office of Naval Research, 1973.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Overall, J.E. & Gorham, D.R. A pattern probability model for the classification of psychiatric patients. Behavioral Science, 1963, 8, 108-116.

Overall, J.E. & Hollister, J.E. Computer procedures for psychiatric classification. Journal of the American Medical Association, 1964, 187, 583-588.

Overall, J.E. & Klett, C.J. Applied multivariate analysis. New York: McGraw-Hill, 1972.

Pecorella, P.A. & Bowers, D.G. Future performance trend indicators: A current value approach to human resources accounting. Technical Report to Office of Naval Research, 1976a.

Pecorella, P.A. & Bowers, D.G. Future performance trend indicators: A current value approach to human resources accounting. Technical Report to Office of Naval Research, 1976b.

Rao, C.R. & Slater, P. Multivariate analysis applied to differences between neurotic groups. British Journal of Psychology 2 (statistical section), 1949, 17-29.

Sashkin, M. Organization development practices. Professional Psychologist, 1973, 4, 187-194.

Sletten, I.W., Altman, H., & Ulett, G.A. Routine diagnosis by computer. American Journal of Psychiatry, 1971, 127, 1147-1152.

Smith, W.G. A model for psychiatric diagnosis. Archives of General Psychiatry, 1966, 14, 521-529.

Spencer, G.J. A methodology for the studies of the impact of organizational values, preferences, and practices on the United States Army. Technical Report to the U.S. Army Research Institute for the Behavioral and Social Sciences, 1975.

Spitzer, R.L. & Endicott, J. DIAGNO: A computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. Archives of General Psychiatry, 1968, 18, 746-756.

Spitzer, R.L., Endicott, J., Cohen, J., & Fleiss, J.L. Constraints on the validity of computer diagnosis. Archives of General Psychiatry, 1974, 31, 197-203.

Taylor, J.C. & Bowers, D.G. Survey of Organizations: A machine-scored standardized questionnaire instrument. Ann Arbor, Michigan: Institute for Social Research, 1972.

Veldman, D.J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart & Winston, 1967.

Wing, J.K. A standard form of psychiatric present state examination and a method for standardizing the classification of symptoms, in Hare, E.H. & J.K. Wing (ed.), Psychiatric epidemiology: An international symposium. London: Oxford University Press, 1970, 93-108.