. . . the diagnosis and understanding of schizophrenia

part II. expanded perspectives for describing and comparing schizophrenic patients

John J. Bartko, John S. Strauss, and William T. Carpenter, Jr.

In part I of this series of reports, we outlined the use of key symptoms as criteria for a diagnosis of schizophrenia. The classification of a patient as schizophrenic on the basis of these symptoms is only the beginning of a more complete description. The next step is to define more broadly the characteristics of patients who have these symptoms and to evaluate from the broader perspective the ways in which they form a single homogeneous group or subgroups, and the ways in which they are similar to or different from other types of patients.

These issues involve consideration of several alternative diagnostic principles. Most fundamental is a decision about the criteria to be used for defining similarity and difference. Will "discriminating symptoms" be the only criteria for determining diagnostic groups? Will other symptoms be given any importance? Will severity of symptoms as well as type of symptoms be included as a criterion for diagnosis? Will the characteristics of the data play a role in establishing diagnostic criteria, or will these criteria be established before the patient group is evaluated? Finally, will diagnostic subgroups be established, and if so, what criteria of similarity and difference will determine the nature of these groups?

These diagnostic issues are all-important clinical issues. The decisions on which criteria to use have important implications regarding whether a particular diagnostic group will have clinically important relationships to etiology, effective treatment, and methods for prevention. There are several methods that can be used in establishing groups according to the different kinds of criteria, for suggesting subgroups to determine in what sense there is a "group of schizophrenias" (Bleuler 1950), and for assigning new patients to established groups.

In this part II, we present methods for describing and comparing patients with discriminating symptoms in terms of a wide range of psychopathologic characteristics, and we describe the assumptions behind these methods, their clinical implications, and the ways they can be used.

Evaluation of Similarities and Differences among Patients

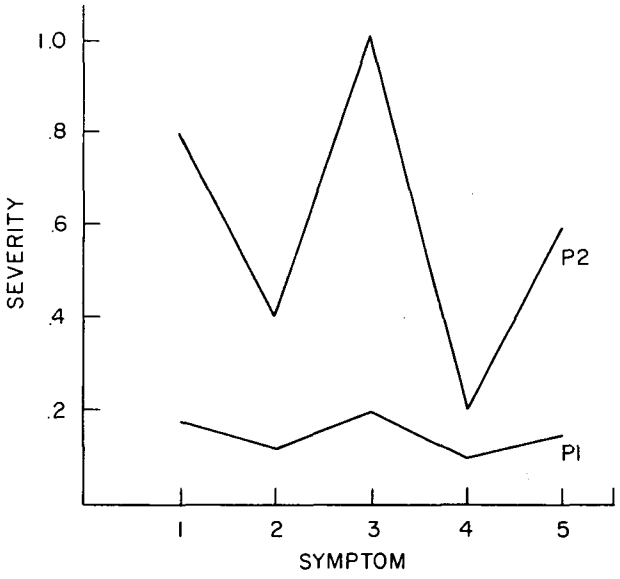
One of the basic tools in making a comparison between persons or groups is the profile—a representation of the individual or group and the degree to which each of a number of characteristics is present. An example of a simple hypothetical profile is presented in table 9 for patients P_1 and P_2, and the plot of the table 9 profiles appears in figure 2.

The profile plots in figure 2 are in rank order (i.e., symptom 3> symptom 1> symptom 5> symptom 2> symptom 4). The profile for P_2 has greater variation and reflects a higher level of severity. These profiles illustrate a basic issue, namely, the existence of alternative criteria for defining similarities and differences among patients.

Table 9. Hypothetical profile of similarities and differences among patients.

Symptom	Percentage of pathology (severity)	
	P_1	P_2
1	17	80
2	12	40
3	20	100
4	10	20
5	15	60

Figure 2. Profile plot for hypothetical data presented in table 9.



If symptom rank order[8] is most important, then the two patients in figure 2 are clinically similar. On the other hand, if severity is more crucial, then the two patients appear to be quite different. In the evaluation of schizophrenia, many psychiatrists would consider the rank order of symptoms more than the severity, in which case, if P_2 is diagnosed schizophrenic, then P_1 would be also.

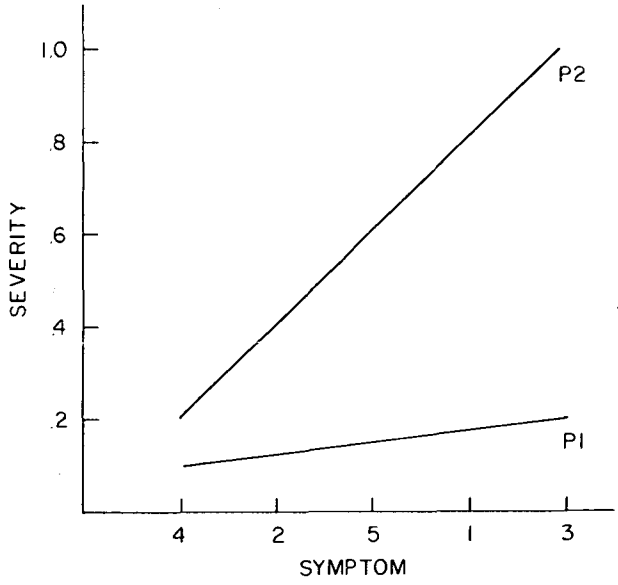
[8] Rank order refers to the way symptoms relate to each other in relative severity. For example, a patient with more depression than paranoia and more paranoia than hallucinations might be called "depressed," whereas a reversal of relative severity might be considered to reflect schizophrenia.

Tau, W, and ANOVA

Common statistical measures for comparing profiles reflect different orientations. The comparison of rank order of symptoms is often assessed by Kendall's tau or, for more than two profiles, by Kendall's coefficient of concordance *W* (Siegel 1956). For the data in figure 2, the patients are entirely similar in the rank order of these symptoms. This profile is depicted more clearly in the rearrangement of variables in figure 3, which also shows, however, that the rank-order notion of similarity does not imply similarity of profiles in a stricter sense.

Profile analysis of variance (ANOVA) is a method that provides a more precise measure of similarity, evaluating not only the rank order of symptom severity for each person but the pattern of symptoms and the level of symptom severity as well (Greenhouse and Geisser 1959). For example, a group of patients with higher scores on hallucinations than depression would not be characterized as similar to a group of schizophrenics unless their hallucination ratings were high in absolute value, not just higher than other symptoms. This technique evaluates whether patients' symptom profiles have similar patterns (parallel) and, if so, whether they have the same severity (level).

Figure 3. Profile plot of figure 2, using table 9 data, with symptoms rearranged.



Profile similarity as defined in the profile ANOVA procedure is illustrated in figure 4. If the distances between profile points at each symptom are equal, as in figure 4, then the profiles are parallel; that is, they have the same pattern.

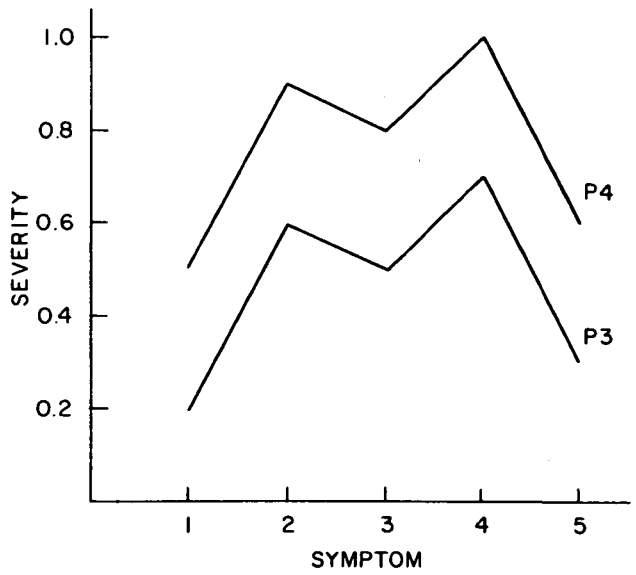
These tau, *W*, and ANOVA approaches provide a method for comparing patients that is more complete than is possible from ratings of presence-absence of a few key symptoms alone. On the other hand, these methods consider each item on the profile as having equal importance, a different approach from most psychiatric theory in which it is assumed that some patient characteristics are more significant than others. In diagnosing schizophrenia, for example, presence of hallucinations is usually weighted more heavily than presence of anxiety. In the discussion of profile comparison, we have assumed that all variables have the same importance. Since this may not be the most productive assumption, it is possible to use other multivariate statistical approaches that permit greater manipulation in determining, comparing, and weighting profiles and variables.

Discriminant Analysis

For some purposes, it is important to determine whether two or more groups of patients have certain more or less consistent characteristics or combinations of characteristics that help to differentiate them. For example, it is of interest to know whether groups of patients diagnosed paranoid schizophrenic consistently have more delusions of persecution, auditory hallucinations, and/or flat affect, or combinations of these symptoms, than patients diagnosed as manic depressive-manic. If the membership of the groups is already known, discriminant analysis (Anderson 1959) can determine which characteristic or combination of characteristics, if any, discriminate among the groups. Weights are assigned to each variable according to its relative importance as part of a combination of variables in discriminating one group of patients from the other. Once discriminating characteristics have been determined, this technique can also be used to assign an individual to one or the other of the groups. The most important characteristics in discriminating the groups are those weighted most heavily in deciding where a patient belongs.

This technique differs from the profile analyses described above in that it evaluates each of the variables to determine which of them is most important in discriminating among known groups. In this sense, it more closely approximates the differential weighting of patient characteristics often used in psychiatric diagnosis than does the unweighted comparison of profiles.

Figure 4. Profile of symptom similarity as defined in an ANOVA procedure.

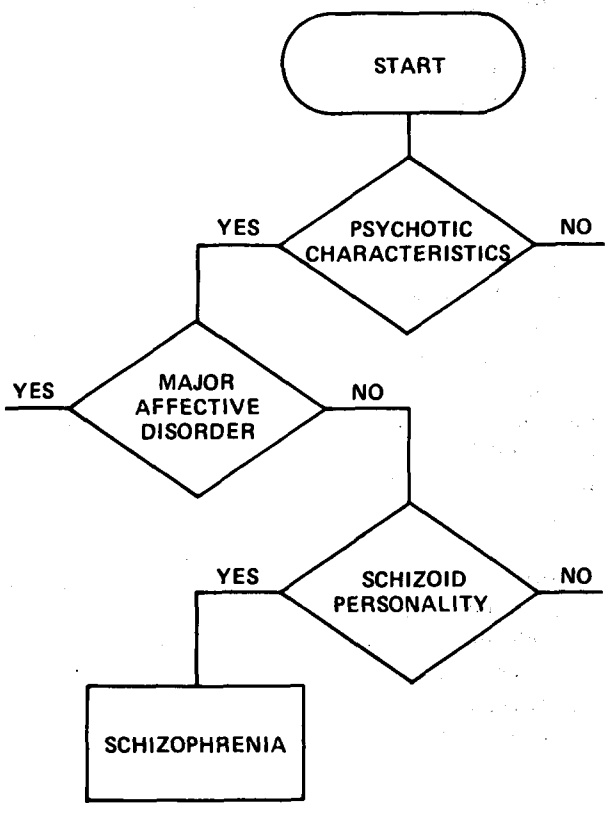


Logical Tree Models

The logical tree approach is another technique for establishing diagnostic groups of patients. This approach has been used in attempts to replicate clinical diagnostic processes by establishing clear, consecutive diagnostic rules (Spitzer and Endicott 1969 and World Health Organization 1973). These decision trees are based on some a priori concept of decision points in reaching established diagnostic groups. DIAGNO (Spitzer and Endicott 1969) uses DSM II (American Psychiatric Association 1968) criteria in defining steps to the schizophrenia branch, while CATEGO (Wing 1973) uses primarily Kurt Schneider's first-rank symptoms. The logical tree method, like discriminant function analysis, is especially suitable for assigning membership of an individual to one of several alternative groups.

The logical tree technique weights variables according to their importance by using some variables at earlier, more crucial, choice points than others. A simple example for the part of such a tree that might be used to establish the diagnosis of schizophrenia is illustrated in figure 5. In this approach, if a patient has psychotic

Figure 5. The logical tree model for establishing diagnostic groups.



characteristics, he will be split off to a given group. All members of that group are then evaluated for major affective disorder. If the patient has no major affective disorder, then he will pass to the next branch that separates patients according to schizoid personality. If he has a schizoid personality, he would remain in the group. Having taken each of the consecutive branching points as he did, he could then be given a diagnosis of schizophrenia. The use of logical trees is most readily carried out with the aid of computer programs. Unlike conventional clinical diagnosis, such programs, in replication, will always yield the same diagnosis in a patient previously passing through the branching series, as well as for any other patients with similar characteristics. In this sense, these types of procedures are 100 percent reliable.

Cluster Analysis

A mathematical formula (algorithm) is used in cluster analysis techniques that compares measurements of patient characteristics (i.e., profiles) to arrive at clusters. These clusters, or "natural groups," are assumed by some investigators to represent those groups of individuals who are most likely to have the largest number of characteristics in common, even those not entered in the algorithm (Sokal and Sneath 1963).

Cluster analysis begins by describing a rule for defining similarity between patients; for example, a rule that patients should be considered similar who have similar shape profiles rather than similar level or rank-order profiles. Once these "natural groups" have been produced, it is possible to define the characteristic profile of each of the clusters and then to develop methods for assigning new patients to the appropriate cluster. The principles for applying cluster analysis to psychiatric data are, however, just beginning to be clarified (Bartko, Strauss, and Carpenter 1971, Fleiss and Zubin 1969, Strauss, Bartko, and Carpenter 1973, and World Health Organization 1973). Several clusters emerged from these studies that had similarities to conventional concepts of schizophrenia but that had clinically distinctive characteristics as well.

Summary of Evaluation Methods

We have discussed alternative models for describing patient similarities and differences and mathematical methods to determine them. Profiles and profile plots lend themselves to graphic communication, thus revealing succinctly what words often obscure. Comparisons of rank-order similarity of profiles are provided by use of the tau and *W* statistics. Analysis of variance permits a more precise comparison of shapes and severity of profiles. When group membership has already been determined, discriminant function analysis assigns weights to the variables that most discriminate the groups. It can also be used in assigning new members. A priori concepts of variables considered crucial in reaching a diagnosis are used in logical tree models, which take certain patient characteristics and assign them levels of priority in determining successive branchings for classifying individuals into diagnostic groups. Cluster analytic techniques search for "natural groups" in a population of patients in which the most meaningful classification is not known. From the assessment of profile similarity to the search for "natural groups" of patients, a considerable range of statistical and investigative tools is available for determining which patients form the most meaningful groups and subgroups.

Table 10. Archetypal data: Profiles of diagnostic groups.[1]

Symptom	Group 1	Group 2	Group 3	Group 4
Confused, vague	0.0	1.6	0.0	0.0
Somatic symptoms	0.0	0.0	0.0	1.1
Withdrawal	0.0	0.9	0.0	1.1
Depressed mood	0.0	0.0	0.0	1.6
Elated mood	0.0	0.0	1.6	0.0
Characteristic hallucinations	1.1	0.0	0.0	0.0
Disordered thoughts	0.9	0.0	0.8	0.0
Loose associations	1.1	0.0	0.0	0.0
Delusions of control	1.6	0.0	0.0	0.0
Limited insight	0.9	0.0	1.1	0.0
Unclear communication	0.0	1.1	0.9	0.0
Rapport problems	0.9	1.2	0.9	0.0
Group size	20.0	20.0	20.0	20.0

[1]Code for symptom ratings: 0 (symptom absent); .5 (presence of symptom uncertain); 1 (symptom definitely present but not continuous or severe); and 2 (symptom definitely present, continuous, and/or severe).

Application with Archetypal Data

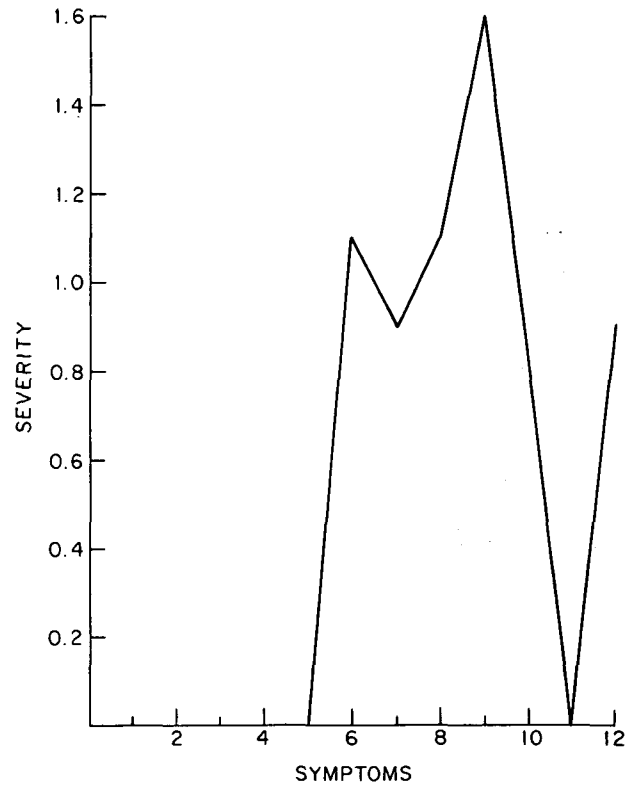
To demonstrate the concepts and techniques described above, a set of data was constructed to represent "patients" who were characteristic of archetypes of different diagnostic categories (Bartko, Strauss, and Carpenter 1971). Four diagnostic groups—G_1, G_2, G_3, and G_4, with 20 patients in each group—were constructed, each with a profile of ratings on 12 symptoms.[9] Each symptom was graded as follows: 0 (symptom absent), .5 (presence of symptom uncertain), 1 (symptom definitely present, but not continuous or severe), and 2 (symptom definitely present, continuous, and/or severe). Table 10 lists the group profiles of the four diagnostic groups and symptom titles. Figures 6, 7, 8, and 9 illustrate the profile plots for the four diagnostic groups.

W and ANOVA

Kendall's coefficient of concordance (W) was computed for the four diagnostic groups to evaluate whether, in terms of symptom rank order, their profiles were similar. The result ($W = .06$) indicated that they were

[9]These four groups may be called paranoid schizophrenia (G_1), simple schizophrenia (G_2), mania (G_3), and psychotic depression (G_4); it should be kept in mind that they are oversimplified archetypes.

Figure 6. Profile plot for paranoid schizophrenia group (G_1).



not. ANOVA was carried out to measure whether the diagnostic groups had similar profile patterns. They did not. Thus, the *W* and ANOVA results agree in confirming four distinct groups. This result is not always the case, as can be seen in the next section on analyzing real data.

Discriminant Analysis

Discriminant function analysis for two groups yields a linear function of the profile variables, which maximally separates the two groups. For illustrative purposes, G_1 and G_2 were selected for discriminant analysis. Variable 9, delusions of control, which was present in G_1 but absent in G_2, entered first in the stepwise program used. This variable provides for 100 percent correct classification of the G_1 and G_2 patients. Such an immediate and complete discrimination is rare with real data. More frequently, a single variable is only the beginning of the discriminating process, which leaves overlap between groups even after many variables are used together.

Figure 7. Profile plot for simple schizophrenia group (G_2).

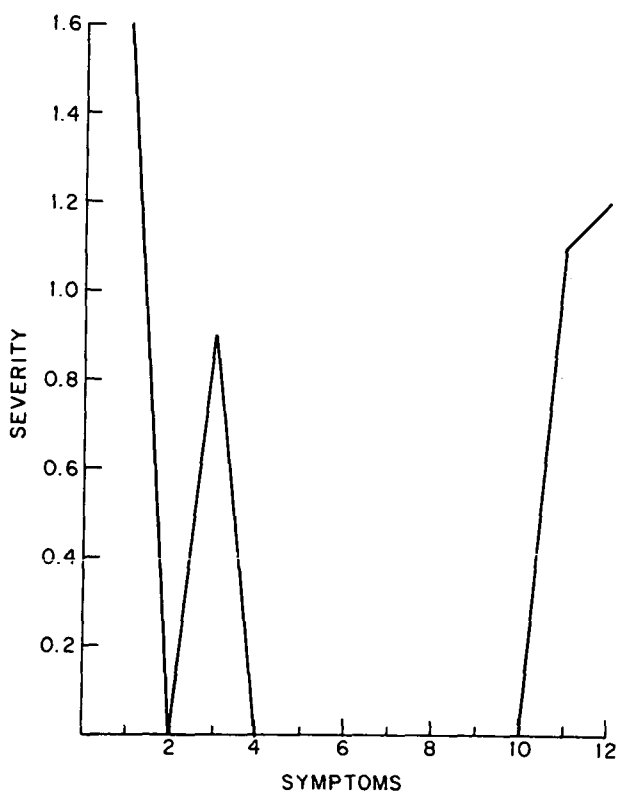
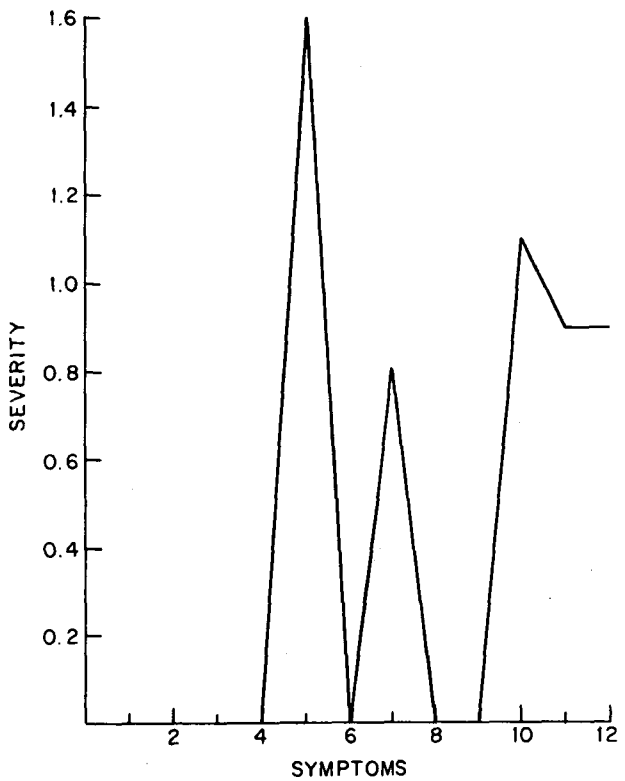


Figure 8. Profile for mania group (G_3).



Logical Tree Model

The logical tree model illustrated in figure 10 was developed to separate the four diagnostic groups. The fashion in which the data were constructed provided for a unique determination of the four diagnostic groups; therefore, the "other" category in figure 10 is not relevant, but is included here only to illustrate a "complete" tree. The path through the tree is evident. If the patient has delusions of control, he is classified G_1; if he has no delusions of control but is confused and vague, then he is classified G_2. If he has no delusions of control, is not confused and vague, but has elated moods, then he is classified G_3. If he has none of the above but does have depressed moods, then he is classified G_4. When "placed in the hopper," these archetypal patients sort themselves into the groups in which they were constructed.

Cluster Analysis

In previous work on cluster analysis (Bartko, Strauss, and Carpenter 1971 and Strauss, Bartko, and Carpenter

1973), one cluster program was found to be the most useful in producing meaningful groups with the data set being analyzed. This program, when used on the archetypal data, reproduced the four diagnostic groups exactly (table 11).

Application with Real Data

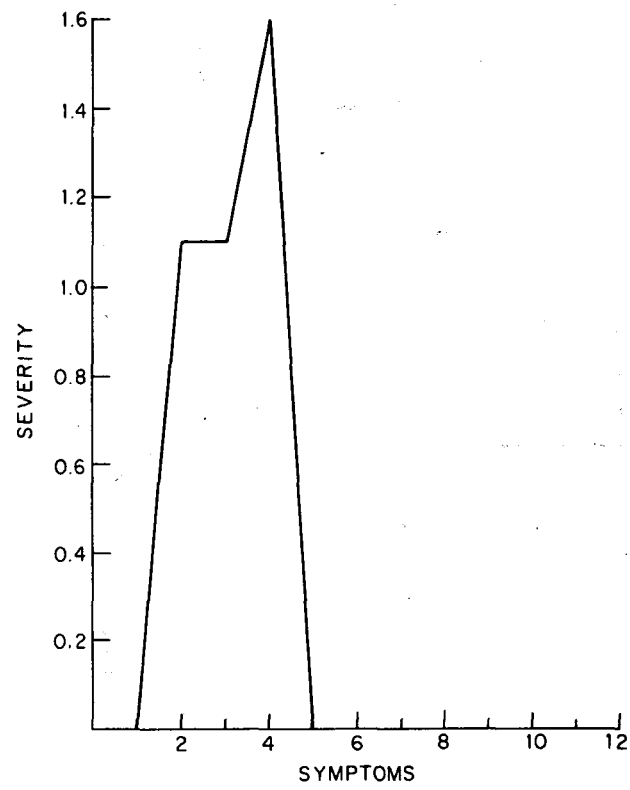
In part I of this series of reports, we described the development of 12 discriminating symptoms for the diagnosis of schizophrenia. Although these symptoms provide an important means of classifying patients, it is possible to go beyond them to illustrate the use of the methods discussed in this part II for deriving a more comprehensive description and comparison of patients. To accomplish this, data on patients evaluated as part of our participation in the IPSS (World Health Organization 1973) were analyzed.

Patients whose profiles are described here are those 97 patients who had been diagnosed as schizophrenic by the authors out of the total cohort of 131 patients evaluated in the U.S. Center of the IPSS. Sixty-six of the 97 patients had six or more of the discriminating symptoms. This group is designated as S_{6+}. The remaining 31 patients (designated S_{5-}) had five or fewer discriminating symptoms. Profiles of these two groups on 11 important symptom dimensions other than the 12 discriminating symptoms appear in table 12, and the profile plots appear in figure 11.

Profile Analysis of Variance and Tau

An ANOVA revealed that profiles for the two groups on the 11 symptoms were not parallel ($p < .05$). Figure 11 also shows that the S_{6+} group had a higher level (more severely symptomatic) than the S_{5-} group.

Figure 9. Profile plot for psychotic depression group (G_4).

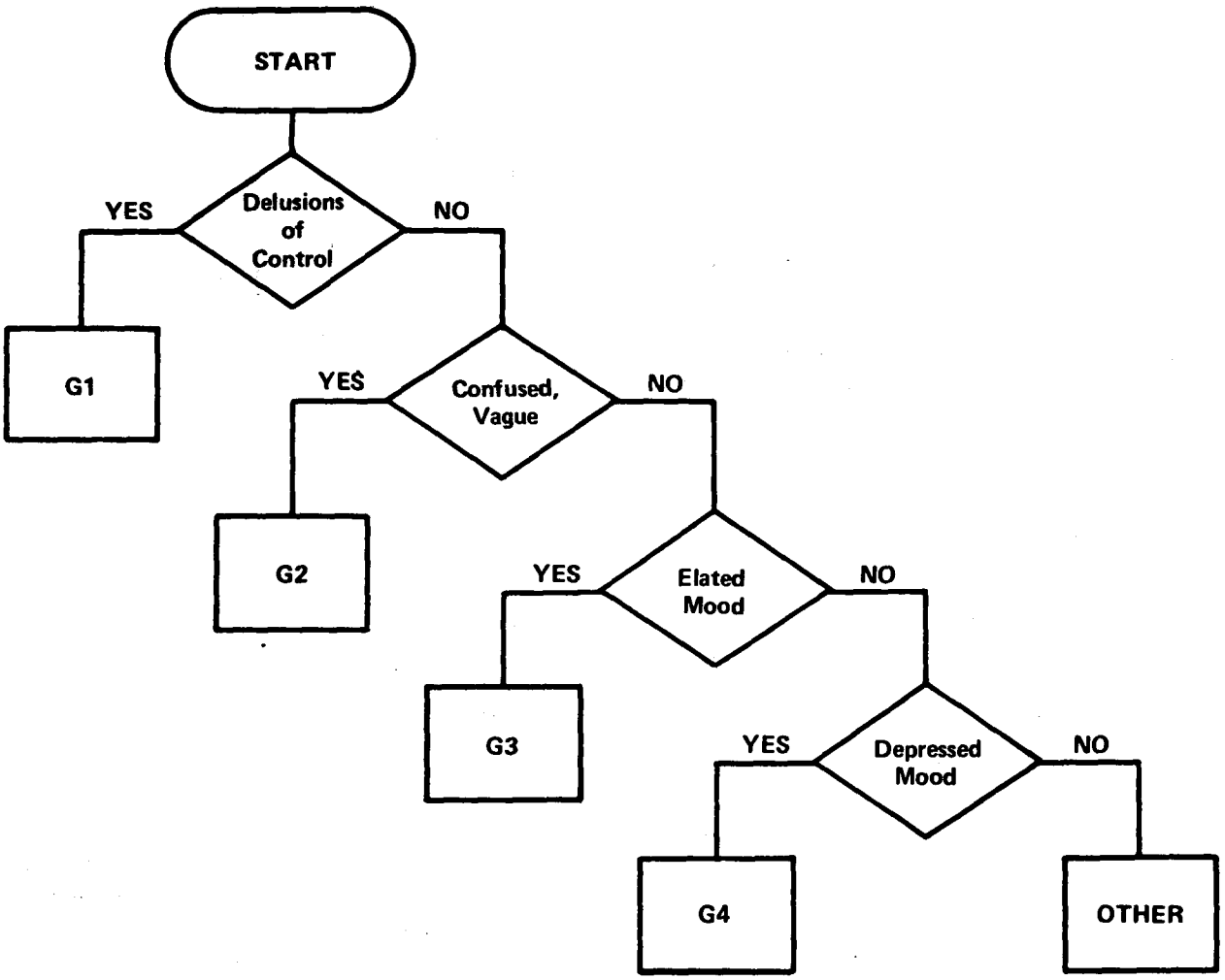


The dimension on which the two groups most differed was paranoid delusions (see table 12). The two groups also differed (t test significant at the $p < .02$ level) on the dimensions of delusions of passivity, labile affect, withdrawal, grandiose delusions, and hallucinations. The remaining five dimensions (depersonalization-derealization, anxiety, retarded speech and/or move-

Table 11. Cluster analysis results.

Diagnostic group	Cluster				
	1	2	3	4	Total
G_1	20				20
G_2		20			20
G_3			20		20
G_4				20	20
Total	20	20	20	20	80

Figure 10. The logical tree model.



ment, bizarre behavior, and nonsocial speech) were not significantly different in the two groups.

This comparison of several important symptoms besides the original discriminating ones provides a more complete picture of these patients and their similarities.

The tau rank-order correlation coefficient on the profile data in table 12 and figure 10 is .51 ($p < .05$). This result, signifying that the profiles are similar by rank order, is in direct conflict with the more precise ANOVA findings signifying nonsimilarity in terms of parallelism.

Discriminant Function

The discriminant function results appear in table 13. Five of the 11 symptom dimensions account for this assignment. The remaining symptom dimensions did not contribute significantly to the classification process. The five symptoms in order of importance are withdrawal, retarded speech and/or movement, nonsocial speech, hallucinations, and depersonalization-derealization. The S_{6+} group exhibited greater pathology for each of these five dimensions, and the percentage of misclassification for the S_{6+} group (14 percent) was less than that for the S_{5-} group (29 percent).

These findings suggest that symptoms other than the original 12 discriminating symptoms are also useful for separating the two groups of patients. Those additional symptoms—withdrawal, retarded speech and/or movement, nonsocial speech, hallucinations, and depersonalization-derealization—provide a more comprehensive pic-

Table 12. Profiles of two groups of schizophrenic patients on 11 symptom dimensions.

Symptom dimensions	Dimension mean scores of patient groups[1]		
	S_{6+}	S_{5-}	*p* value (*t* test)
Paranoid delusions	.321	.092	.001
Delusions of passivity	.225	.046	.001
Labile affect	.174	.032	.001
Withdrawal	.217	.089	.001
Grandiose delusions	.153	.049	.001
Depersonalization-derealization	.205	.112	.07
Hallucinations	.150	.074	.02
Anxiety	.214	.180	.17
Retarded speech and/or movement	.077	.051	.17
Bizarre behavior	.025	.014	.14
Nonsocial speech	.011	.002	.14

[1] S_{6+} indicates 66 NIMH diagnosed schizophrenic patients with six or more differential symptoms; and S_{5-} indicates 31 NIMH diagnosed schizophrenic patients with five or fewer differential symptoms. Dimension scores range from 0 (no evidence of symptoms) to 1 (maximum symptomatology).

Figure 11. Profile plots of schizophrenic groups with six or more or five or fewer discriminating symptoms.

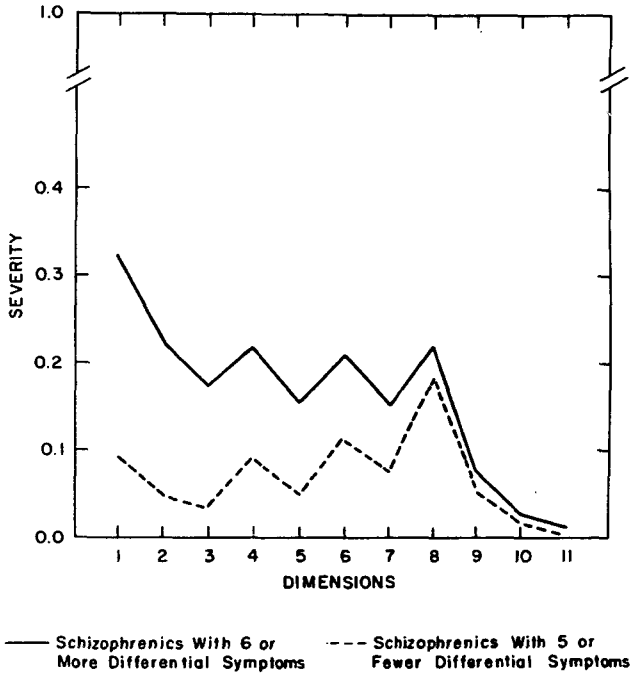


Table 13. Discriminant function assignment on 11 symptom dimensions.

Patient group[1]	Discriminant function assignment		Total
	S_{6+}	S_{5-}	
S_{6+}	57	9	66
S_{5-}	9	22	31
Total	66	31	97

[1] S_{6+} indicates 66 NIMH diagnosed schizophrenic patients with six or more differential symptoms; and S_{5-} indicates 31 NIMH diagnosed schizophrenic patients with five or fewer differential symptoms.

ture of the distinguishing characteristics of the two groups than the 12 discriminating symptoms alone.

Logical Tree Analysis

Results of a CATEGO analysis (Wing 1973) of the 97 patients can be found in table 14. Using only the 11 symptoms shown in this table, a significant relationship

Table 14. Analysis of three classes of patients on 11 symptom dimensions.

Patient group[1]	CATEGO class: Schizophrenic and paranoid psychoses (class 1)	Manic psychoses (class 2)	Depressive psychoses and neuroses (class 3)	Total
S_{6+}	53	10	3	66
S_{5-}	16	4	11	31
Total	69	14	14	97

[1] S_{6+} indicates 66 NIMH diagnosed schizophrenic patients with six or more differential symptoms; and S_{5-} indicates 31 NIMH diagnosed schizophrenic patients with five or fewer differential symptoms.

was found ($\chi^2 = 16.5$, $df = 2$, $p < .001$) between the two systems of classification, S_{6+}/S_{5-}, and CATEGO. The similarity is a limited one, however; the major contributor of the chi-square value is from only one CATEGO class—namely, patients in class 3, the depressed class.

Cluster Analysis

Results of a two-group cluster analysis can be found in table 15. In this analysis, the cluster technique, using the 11 added symptoms, did not sort out the patients with respect to their S_{6+} versus S_{5-} membership ($\chi^2 = 1.6$, $df = 1$, $p > .10$). When the 11 symptoms, all weighted equally, are used as criteria, the S_{6+} and S_{5-} groups of patients are lost with the cluster technique used, and two groups with other characteristics are identified.

Table 15. Cluster analysis of two groups on 11 symptom dimensions.

Patient group[1]	Cluster 1	Cluster 2	Total
S_{6+}	50	16	66
S_{5-}	27	4	31
Total	77	20	97

[1] S_{6+} indicates 66 NIMH diagnosed schizophrenic patients with six or more differential symptoms; and S_{5-} indicates 31 NIMH diagnosed schizophrenic patients with five or fewer differential symptoms.

The different methods described above for comparing patients enable consideration of a large number of variables in determining patient groups. They also demonstrate the importance of evaluating which one of the alternative measures of similarity is most valid and what its implications are in determining, for clinical or research use, why a given group of schizophrenics should be considered together.

Conclusions (Part II)

Diagnosing and comparing schizophrenic patients in terms of a large number of characteristics can be carried out by several basically different methods. Each method represents a conceptual model that varies from the others along several important parameters. Some approaches establish a priori groups based on theoretical or other conceptions, whereas other approaches permit the data to influence the kinds of diagnostic groups constructed. Some approaches depend more on rank order of symptoms in comparing and constructing groups, whereas others concentrate more on patterns of symptomatology or severity of symptoms or both. Some approaches weight characteristics differentially, whereas other approaches consider all characteristics of similar importance.

The different methods do not provide an answer to the question, What is the best perspective for viewing

patients with the discriminating symptoms of schizophrenia? There is no single answer. Instead, there are several alternative approaches, each with a different conceptual basis and each with certain fundamental assumptions about what parameters are crucial in determining when patients form a homogeneous group. The most useful of these alternative perspectives remains to be determined. This process of deciding which model is best will have to be accomplished by demonstrating key relationships of the particular model to the major clinical criteria, etiology, response to treatment, and outcome.

The Authors

John J. Bartko, Ph.D., is Mathematical Statistician, Theoretical Statistics and Mathematics Branch, National Institute of Mental Health, Bethesda, Md.; John S. Strauss, M.D., is Director, Clinical Psychiatry Research Programs, University of Rochester School of Medicine and Dentistry, Rochester, N.Y.; and William T. Carpenter, Jr., M.D., is Acting Chief, Section on Psychiatric Assessment, NIMH, Bethesda, Md.

an invitation to readers

Providing a forum for a lively exchange of ideas ranks high among the *Schizophrenia Bulletin's* objectives. In the section, **At Issue,** readers are asked to comment on specific controversial subjects that merit wide discussion. But remarks need not be confined to the issues we have identified. **At Issue** is open to any schizophrenia-related topic that needs airing. It is a place for readers to discuss articles that appear in the *Bulletin* or elsewhere in the professional literature, to report informally on experiences in the clinic, laboratory, or community, and to share ideas—including those that might seem to be radical notions. We welcome all comments.—*The Editors.*

Send your remarks to:

At Issue
Center for Studies of Schizophrenia
National Institute of Mental Health
5600 Fishers Lane
Rockville, Md. 20852