A MODEL FOR MEASURING THE CONSISTENCY OF DIAGNOSTIC CLASSIFICATION

S. FICHMAN
D. EDWARDS

REPORT NO. 76-47

See 1473 in deck.
A Model for Measuring the Consistency of Diagnostic Classification

Susan Fichman and Darrel Edwards
Naval Health Research Center
San Diego, California

Report No. 76-47. The views presented in this paper are those of the authors.
No endorsement by the Department of the Navy has been given or should be inferred.
A Model for Measuring the Consistency of Diagnostic Classification

Susan Fichman and Darrel Edwards

One goal of clinical practice is to develop a rational system for making and using a diagnosis in treatment. During clinical practice, clinicians develop decision making rules for evaluating their clients. These rules will guide future interactions. If the rules are sound, the clinician will be able to successfully deal with his clients. The measurement of the actual accessibility of these cognitive rules to the trainee in managing clinical cases is difficult. When experimenters rely on observable behaviors to evaluate clinical functioning, they ignore the fact that different clinical styles may be equally effective in dealing with a patient. Observable clinical behaviors may vary widely. There are alternative decision making processes which lead to successful treatment outcomes (Edwards, Gunderson, Brown, and Taylor, 1973).

This study examined the conceptual process of diagnosis as used by psychiatrists in their clinical practice. Psychiatrists evaluated their patients using dimensions unique to their clinical philosophy. These evaluations were analyzed to produce each psychiatrist's conceptual scheme as it relates to the diagnosis of his patients. The purpose of this study was to demonstrate that the diagnostic structure of an individual clinician can be objectively measured, producing a method for evaluating consistency, clarity, or other dimensions of the clinical decision.
A grid method (Bannister and Mair, 1968) was used to collect data from 13 psychiatrists at two Navy inpatient psychiatric services. The doctors had a mean time in practice post-residency of 2.0 years with a range of 1 to 12 years.

Grids were administered in small groups of no more than five clinicians at one time. The clinicians were asked to compile a list of 16 patients from their caseloads: four psychotics, three neurotics, three personality disorders, three situational maladjustments, and three alcoholics. Adjustments in the distribution were made to accommodate differences in individual caseloads where a clinician was not able to fill all classes. Adjustments were made by having the therapists substitute patients from other diagnostic categories when they had exhausted their cases from the requested category. At least two patients from each diagnostic group were required.

Each clinician was instructed to evaluate his patients on 16 clinical dimensions which the therapist chose as important in his practice. Patients were scored with an "X" to indicate whether that patient possessed a specific characteristic. This technique produced a 16 dimensional profile for each patient. The result was a grid of 16 patients by 16 dimensions (Figure 1).

Data Analysis

The intention of the analysis was to determine if a clinician exhibited a conceptual scheme which indicated that similarly diagnosed cases were treated with some consistency. Each grid was analyzed individually with the following
<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>CONTRAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Honest</td>
<td>Dishonest</td>
</tr>
<tr>
<td>2 Mature</td>
<td>Immature</td>
</tr>
<tr>
<td>3 Insight</td>
<td>No Insight</td>
</tr>
<tr>
<td>4 Acceptance of Illness</td>
<td>Denial of Illness</td>
</tr>
<tr>
<td>5 Considerate</td>
<td>Egocentric</td>
</tr>
<tr>
<td>6 Independent</td>
<td>Dependence</td>
</tr>
<tr>
<td>7 Self-conscious</td>
<td>Self-assured</td>
</tr>
<tr>
<td>8 Hyper-active</td>
<td>Psycho-motor retarded</td>
</tr>
<tr>
<td>9 Many friends</td>
<td>Improvised Interpersonal relationships</td>
</tr>
<tr>
<td>10 Oral Character</td>
<td>Not Oral Character</td>
</tr>
<tr>
<td>11 Obsessive Compulsive</td>
<td>Not Obsessive Compulsive</td>
</tr>
<tr>
<td>12 Manipulative</td>
<td>Not Manipulative</td>
</tr>
<tr>
<td>13 Poor Impulse Control</td>
<td>Good Impulse Control</td>
</tr>
<tr>
<td>14 Poor Ego Image</td>
<td>Good Ego Image</td>
</tr>
<tr>
<td>15 Covert Hostility</td>
<td>No Covert Hostility</td>
</tr>
<tr>
<td>16 Lack of Energy</td>
<td>Adequate Energy</td>
</tr>
</tbody>
</table>
procedure. Every patient's 16-dimensional profile was compared with every other patient's profile for each clinician's grid. Paired profiles were examined to determine the number of matches in dimensions ascribed to the pairs of patients. A match occurred when two patients were rated as having the same attribute in common or when both patients did not possess a given attribute. Profiles did not match when one patient was rated as having an attribute and the other patient was not. The number of matches was a measure of similarity between patients. The higher the total, the greater the perceived similarity.

The similarity totals produced a new patient profile listing the similarity between that patient and every other patient. This information was recorded in a 16 by 16 matrix. Intercolumnar correlations were computed, i.e., each patient's similarity profile was correlated with every other patient's similarity profile. A cluster analysis (Johnson, 1967) was performed on these correlations. Patients whose profiles were highly correlated were combined into clusters. The program scanned the matrix of correlations and selected the highest correlation. The similarity score profiles for the two most highly correlated patients were collapsed to form a single profile. Correlations were recomputed to form a new matrix containing one fewer pair of correlations and the new correlation matrix was scanned to again select the highest correlation. This process was stopped after the tenth reiteration to maximize the possibility of grouping patients into five clusters representing the five diagnostic categories.

A cluster analysis was performed on each clinician's data. Four representative cluster patterns are presented in Figures 2-5. The patient's diagnosis is represented by an abbreviation (psychosis (P), neurosis (N), situational
maladjustment (SM), personality disorder (PD), and alcoholism (A), the patient's number is a subscript. The relative lengths of the "branches" represent the correlation for the cluster; longer distances represent lower correlations.

Results and Discussion

Clinician I. The first cluster (r=1.0) was a pair of situational maladjustments. At later stages, clusters contained combinations of unlike diagnoses. One alcoholic defined a single member case. Final clusters showed some appropriate pairings of diagnoses but little overall structure for diagnostic clarity.

Figure 2. Diagnostic Structure for Clinician I.
Clinician 2. The initial clusters ($r > .90$) showed diagnostic clarity. The first contained two alcoholics, the second contained two psychotics, the third contained two personality disorders, the fourth contained two psychotics, and the fifth contained one psychotic and one neurotic. Later clusterings produced groups of mixed diagnoses. Two personality disorders were single member cases that did not combine with any other cases. This diagnostic structure began with pairings of similar diagnoses but these pairings did not generalize to include more members of the same diagnostic class.

![Figure 3. Diagnostic Structure for Clinician 2.](image-url)
Clinician 3. At early stages ($r > .80$), there is evidence for diagnostic clarity with one cluster containing two psychotics, a second cluster containing two neurotics, a third containing a psychotic and an alcoholic, and a fourth containing a psychotic and a personality disorder. Later stages ($r > .60$) showed increased diagnostic clarity. One cluster contained two of four psychotics, a second cluster contained all neurotics, a third cluster contained all situational maladjustments and one personality disorder, a fourth contained a psychotic and an alcoholic and a fifth contained a psychotic and a personality disorder. One alcoholic and one personality disorder defined single member classes.

Figure 4. Diagnostic Structures for Clinician 3.
Clinician 4. Initial clusters (r > .90) showed a clear diagnostic strategy. Of the five clusters produced, only one cluster contained unmatched diagnoses. Later clusters (r > .970) showed continued diagnostic clarity. One cluster contained all alcoholics, a second cluster contained one psychotic and one situational maladjustment, a third cluster contained all neurotics, a fourth cluster contained two of three situational maladjustments, and a fifth cluster contained all personality disorders. One psychotic defined a single member case. This diagnostic structure used distinct categories for every diagnosis but psychosis.

Figure 5. Diagnostic Structure for Clinician 4.
Summary of all Clinicians

Individual cluster patterns were unique in composition, but there were some similarities in structure across clinicians. The four examples described above represent the four types of structure in diagnosis: (1) General inconsistency; (2) Immediate consistency followed by inconsistency; (3) Developing consistency; and (4) General consistency. Cluster analysis of each profile showed the diagnostic structure of each clinician. A summary of the profiles had the following characteristics: (a) four clinicians exhibited a consistency in conceptually sorting all patients; (b) six clinicians had a consistent structure for one or two of the diagnostic classes with each diagnosis showing some consistency for at least one doctor in the sample; and (c) three clinicians appeared to have no conceptually clear way to manage their patient sample.

Clarity of diagnostic structure appeared often at the concrete level of association and continued to build as associations became more abstract. Most doctors appeared to have some systematic structure associated with diagnosis. The complexity of these structures ranged from single concrete pairings to more complex structures handling one diagnostic category to those structures that dealt efficiently with a wide range of diagnostic classes. Individual diagnostic systems differed in the level of abstraction associated with diagnostic clarity and confusion. These differences were objectified by the grid method used in this research. Overall, the results defined the level of functioning for each clinician in diagnosing patients.

Comment

Although variations in clinical cases and individual structures are expected, some consistency in diagnostic practices is also anticipated. Diagnostic similarity is expected to be greater within diagnostic categories than across
diagnostic categories, even if the diagnostic processes were different for each clinician. Most doctors in this study demonstrated clear, consistent conceptual structure for the diagnostic categories.

This study clearly demonstrates that clinical decision making processes can be documented. Documentation of a subjective process would permit comparisons between doctors, between doctors and an idealized standard, and between teachers and students in clinical training settings. Objectifying diagnostic decisions allows a clearer communication between supervisor or teacher and trainee about clinical rules and subjective processes (Ivery, et al., 1968). Previous use of introspective processes and behavioral descriptions of diagnosis have failed to delimit the dynamics of the process (Hansen and Barker, 1964).

Information on decision making would be useful in a training situation. In an ongoing training program, both the instructor and the student could complete grids on patients familiar to both. The instructor could compare the student's structure and his own and evaluate differences to determine where the student might need additional feedback. Thus, the instructor could spot student weaknesses and direct his attention toward these areas. Repeated examination, comparing grids produced by students and teachers at additional checkpoints, would show the students' progress.

Program evaluation may also be measured with these procedures. Selected observation points might be chosen and students evaluated on the critical dimensions to be taught in a training program. The results could be compared to benchmarks for the program. Effectiveness of training methods, trainers, or program material may be reflected against criteria developed by the documentation methods described in this report (Carkhuff & Truax, 1965). This technique provides a method for measurement which is relatively easy to use and shows possibilities for criteria development. Incorporating cognitive information with current behavioral observations will give a more complete picture of clinical functioning.
A Model for

References


A Model for Measuring the Consistency of Diagnostic Classification

Susan Fichman and Darrel Edwards

Naval Health Research Center
P.O. Box 85122
San Diego, CA 92138

Naval Medical Research and Development Command
Bethesda, Maryland 20014

Bureau of Medicine and Surgery
Department of Navy
Washington, D.C. 20375

Approved for public release, distribution unlimited.
psychiatric environment. Evaluation of clinical training is a complex task. Traditional methods have relied on directly observable criteria and often produced disappointing results. The real object of clinical training is to develop good decision making procedures. This paper presents a methodology for examining the conceptual process of diagnosis. Thirteen psychiatrists evaluated their own patients using a grid method. Each doctor's grid was examined individually using cluster analysis. The Doctor's cluster structures were evaluated to determine whether the psychiatrists were systematically using diagnosis to classify the patients. Application of this technique to the training setting was presented.