

# DIAGNO

**A Computer Program for Psychiatric Diagnosis  
Utilizing the Differential Diagnostic Procedure**

*Robert L. Spitzer, MD, and Jean Endicott, PhD, New York*

# DIAGNO

## A Computer Program for Psychiatric Diagnosis Utilizing the Differential Diagnostic Procedure

*Robert L. Spitzer, MD, and Jean Endicott, PhD, New York*

**I**N RECENT YEARS there has been increased interest in using computers to arrive at a clinical diagnosis. Part of the present well-documented unreliability of psychiatric diagnoses<sup>1,2</sup> lies in the variability in the operations by which clinicians use the raw data of observation to make a diagnosis. This source of unreliability is completely eliminated by the use of a computer program which will always arrive at the same diagnosis when given the raw data describing a subject. The availability of a computer program for psychiatric diagnosis with demonstrated validity would make possible meaningful comparisons of the diagnostic composition of various populations. Such comparisons are now difficult to interpret because of the use of different diagnostic criteria by clinicians. Although several attempts at developing computer programs for classifying patients according to the standard psychiatric nomenclature have been made, they have all relied on various statistical models based on either the probability of various symptom patterns occurring in the

various diagnoses or statistical procedures based on measures of profile similarity.<sup>3-6</sup>

This paper describes a computer program for psychiatric diagnosis, DIAGNO, which is based not on a statistical model but rather on a logical decision tree model similar to the differential diagnostic procedure employed in clinical medicine. The decision tree approach consists of a series of questions, each of which is either true or false. The result of each question rules out one or more diagnoses or groups of diagnoses and determines the next question asked. One advantage of this model over statistical models is that it does not require knowledge or estimates of the base rate of occurrence of symptoms or signs for each diagnosis (which is likely to vary from population to population). Furthermore, since the clinical criteria which determine the decisions are understandable to clinicians, it is possible to subject the decision tree program to clinical scrutiny so that modifications may be made to more closely approximate clinical practice.

Submitted for publication Nov 6, 1967.

From the Research Division, Washington Heights Community Service, New York State Psychiatric Institute and Biometrics Research, New York State Department of Mental Hygiene, New York.

Read before the annual meeting of the American Psychiatric Association, Detroit, May 8-12, 1967.

Reprint requests to 722 W 168th St, New York 10032 (Dr. Spitzer).

### Development of Program

The input data for DIAGNO are the information on current psychopathology, contained in 39 scale scores (Table 1) of the Psychiatric Status Schedule, (R. L. Spitzer, et al, *The Psychiatric Status Schedule: Tech-*

nique for Evaluating Social and Role Functioning and Mental Status, unpublished paper), as well as age, sex, and number of previous psychiatric hospitalizations. The Psychiatric Status Schedule (PSS) is an instrument which contains both a standardized interview schedule and a matching inventory of 492 precoded items (Table 2) descriptive of small units of psychopathology. The interviewer uses the interview schedule to elicit information needed to judge the items of the inventory. The focus is upon the subject's functioning during the past week. The evaluation takes from 45 to 75 minutes. Except where noted, all PSS evaluations reported in this paper were done by research interviewers whose backgrounds included training in psychology, social work, and two years of medical school.

The output of DIAGNO is one of 25 standard American Psychiatric Association (APA) diagnoses and qualifying phrases,<sup>7</sup> as well as two unofficial diagnoses: not ill and nonspecific illness with mild symptomatology (Table 3). These standard APA diagnoses accounted for 88% of diagnoses assigned to diagnosed psychiatric patients admitted to Bellevue Psychiatric Hospital during 1964.

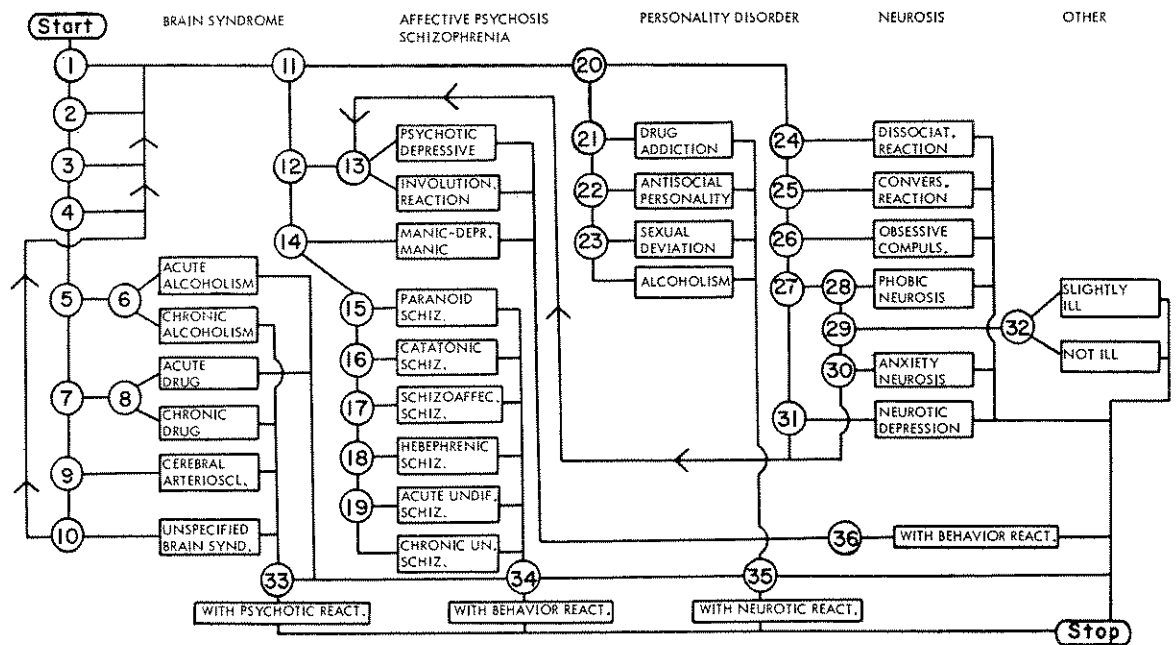
Since the PSS is limited to current status, diagnoses relying heavily on historical infor-

mation, such as some of the personality disorders, are not included. Other diagnostic categories are combined. For example, manic depressive psychosis, depressed type, is subsumed under psychotic depressive reaction, and senile brain disease is subsumed under cerebral arteriosclerosis. As information is not available on the PSS relative to psychophysiological disorders or mental deficiency, these groups are not included.

The first step in the development of DIAGNO was to schematize in a flow chart (Figure) the basic logic implicit in the APA nomenclature. Thus, the first consideration is whether or not there are signs suggestive of an organic brain disorder. The next decision is whether there are signs suggestive of a psychotic disorder. Then the sociopathic personality disorders are considered, followed by the neuroses.

The next step was to define each decision in terms of raw scale score cutoff points and logical operators. Some decisions are simple. For example, decision one (organicity?) is merely "If Disorientation-Memory greater than 3 is true, go to decision 2; if false, go to decision 11." Other decisions are very complex. For example, decision 11 (psychosis?) is "If (Delusions-Hallucinations greater than 2) or (Retardation-Withdrawal greater than 4) or (Inappropriate-Bizarre Appear-

Schematic flow chart for DIAGNO computer program.



**Table 1.—Psychiatric Status Schedule Scales Used in DIAGNO**

Scale Name	No. of Items
Inappropriate or bizarre appearance or behavior	18
Belligerence-negativism	23
Agitation-excitement	8
Retardation-emotional withdrawal	15
Speech disorganization	13
Suspicion-persecution-hallucinations	24
Grandiosity	7
Depression-anxiety	51
Suicide-self mutilation	9
Somatic concerns	14
Social isolation	13
Daily routine-leisure time impairment	24
Sociopathic impulses or acts	14
Alcohol abuse	18
Drug abuse	23
Disorientation-memory	13
Employed wage earner	17
Housekeeper	13
Student or trainee	25
Total score	362
Anxiety	14
Auditory hallucinations	5
Catatonic behavior	2
Conversion reaction	1
Delusions-hallucinations	28
Depression-suicide	11
Dissociation	3
Elated mood	2
Guilt	7
Obsessions-compulsions	3
Persecutory delusions	7
Phobia	2
Roles	17
Sex deviation	4
Silliness	1
Visual hallucinations	3
Miscellaneous	55
Validity check	3

**Table 2.—Items in the Grandiosity Scale of the Psychiatric Status Schedule**

Item Weight	Item No.	Item Content
1	47	In appraising himself he indicates an inflated view of his value or worth [grandiosity].
1	66	Says he is in love with or loved by everybody or some person he does not know (eg, interviewer).
3	194	Indicates his voices refer to his exalted power, status, or knowledge.
3	370	Claims power or knowledge beyond the bounds of credibility (eg, has special relation to God, can read people's minds) [delusion].
3	374	Assumes the identity of a famous figure or makes impossible claim of personal fame.
1	416	Speaks of contact, power, knowledge, or sensational plan which, though not impossible, is extremely unlikely [nondelusional grandiosity].
1	417	Keeps boasting of his accomplishments, skills, influence, or contacts.

**Table 3.—DIAGNO Output**

Diagnoses
Acute brain syndrome, alcohol intoxication
Acute brain syndrome, drug intoxication
Chronic brain syndrome, alcohol
Chronic brain syndrome, drug
Chronic brain syndrome, cerebral arteriosclerosis
Chronic brain syndrome, unspecified cause
Involuntal psychotic reaction
Manic depressive, manic
Psychotic depressive
Schizophrenic reaction, hebephrenic
Schizophrenic reaction, catatonic
Schizophrenic reaction, paranoid
Schizophrenic reaction, acute undifferentiated
Schizophrenic reaction, chronic undifferentiated
Schizophrenic reaction, schizo-affective
Psychoneurotic reaction, anxiety
Psychoneurotic reaction, dissociative
Psychoneurotic reaction, conversion
Psychoneurotic reaction, phobic
Psychoneurotic reaction, obsessive-compulsive
Psychoneurotic reaction, depressive
Sociopathic personality, antisocial
Sociopathic personality, sexual deviation
Sociopathic personality, alcoholism
Sociopathic personality, drug addiction
Nonspecific illness with mild symptomatology
Not psychiatrically ill
Qualifying Phrases
Without qualifying phrase
With psychotic reaction
With neurotic reaction
With behavioral (personality disorder) reaction

ance or Behavior plus Retardation-Withdrawal plus Speech Disorganization greater than 7) or (Elation greater than 0 or Grandiosity greater than 2) and (Agitation-Excitement greater than 2) or (Speech Disorganization greater than 3) or (Social Isolation greater than 7 and Alcohol Abuse less than 5) is true, go to decision 12; if false, go to decision 20." Note that the implications of some decisions are not irrevocable in that they may be rejected by the results of later decisions. For example, subjects who meet the criteria of decision 1 suggesting an organic brain syndrome are examined at three later decision points, the results of which may exclude this group of diagnoses from consideration (decisions 2, 3, and 4). Similarly, subjects who do not meet the criteria for decision 11, which suggests psychosis, may be examined at two later decision points, the results of which are diagnoses of one of the psychoses (decisions 30 and 31).

Real cases were used in further developing the program. On the basis of clinical assessment, one of the 27 diagnoses was assigned by the authors to each of 140 individ-

ual PSS protocols obtained from a sample of psychiatric admissions. DIAGNO was repeatedly modified until its diagnoses and the clinical diagnoses were the same for each case. The final version of the full program utilizes 36 decisions and is written in FORTRAN IV for the IBM 7094.

### Measure of Diagnostic Agreement

The measure of agreement in the studies reported below is a statistic developed for indexing agreement on nominal scales called kappa.<sup>8</sup> Kappa is the proportion of agreement corrected for, ie, over and above, expected chance agreement, and varies from negative values for less than expected chance agreement through 0 for chance agreement to +1.0 for perfect agreement. A modification of kappa, called weighted kappa, is used when disagreements on a nominal scale are weighted for gravity of disagreement. A slightly modified version of Sandifer's (Degrees of Diagnostic Deviance, unpublished data, January 1966) method for quantifying levels of diagnostic disagreement was used in the computation of weighted kappa. For example, if the pair of diagnoses are in complete agreement, ie, to four digits, the level of disagreement is 0. If the pair is neurotic depressive reaction and psychotic depressive reaction, the level of disagreement is 3. If the pair is neurotic depressive reaction and paranoid schizophrenia, the level of disagreement is the maximum, 9. The rationale and advantages of kappa, weighted kappa, and the use of levels of diagnostic disagreement are described in an earlier paper.<sup>9</sup>

So that the reader can have a standard for kappa against which he can compare the kappas in the studies reported below, it should be noted that in two studies by Sandifer, in which psychiatrists observed the same interviews and had summaries of historical information, the kappas for 31 cases and 36 cases were 0.50 and 0.59 respectively.

The diagnostic agreement in these two studies is reported by Sandifer to be comparable to that found in other studies by him and his associates. (These kappas were calculated on unpublished data supplied to us by Sandifer.)

### Reliability

Of course, the reliability of DIAGNO is perfect when it is repeatedly given the same input. A test of reliability, comparable to the usual test of the reliability of clinical diagnosis, is to compare the DIAGNO diagnoses for two observers who independently evaluate a series of patients, each observer recording his judgments on the PSS.

Eight raters, paired in various combinations, with one of them conducting the PSS interview, independently evaluated 55 newly admitted psychiatric inpatients. The weighted kappa for the 55 pairs was 0.66.

### Validity Studies

Since the aim of DIAGNO is to simulate clinical judgment, its validity was tested in a series of studies by comparing the computer diagnoses with diagnoses made by clinicians on samples of subjects. None of the 42 clinicians who made diagnoses, or the 16 research PSS interviewers who participated in the studies reported in this paper, were in any way familiar with the basic design or decisions of DIAGNO.

In the studies described below, DIAGNO diagnoses were compared with clinical diagnoses made by psychiatrists under three different conditions:

(1) From PSS protocols alone. These protocols consist of a listing of all of the items used by DIAGNO grouped by scale as well as age, sex, and number of previous hospitalizations. The items judged true for a particular subject were checked on the protocol. Here the psychiatrist and DIAGNO were both limited to the PSS items. However, the input for DIAGNO was scale

Table 4.—Chance Corrected Diagnostic Agreement (Weighted Kappa) for 200 PSS Protocols

	Packet*				
	A	B	C	D	E
Clinician 1—					
Clinician 2	0.52	0.38	0.45	0.25	0.25
Clinician 1—					
DIAGNO	0.44	0.20	0.35	0.33	0.14
Clinician 2—					
DIAGNO	0.47	0.15	0.22	0.37	0.13

\*Each packet contained 40 cases.

scores, whereas the psychiatrist was presented with the individual items grouped by scale.

(2) After conducting a PSS interview. Here it was the psychiatrist's judgments of the PSS items which constitute the input for DIAGNO. However, the psychiatrist also had available to him the unscorable aspects of the subject's interaction and behavior, and in some instances, spontaneously elicited historical information.

(3) After the usual hospital admission work-up and admission conference. Here the psychiatrist had no knowledge of the PSS

evaluation and was not limited to the patient as a source of information. Furthermore, historical information was virtually always available.

The first study was designed to test DIAGNO with all of the 27 possible diagnoses, since many were not present in the 140 real cases used in the initial development of the program. A second purpose was to compare the ability of DIAGNO with that of a clinician to recognize the 27 diagnoses. To do this hypothetical cases were used. A group of 27 psychiatric residents (predominantly second and third year) and three attending psychiatrists each completed PSS protocols noting the items true, as well as the age, sex, and number of previous hospitalizations, for hypothetical patients with one of the 27 diagnoses. Each of the 30 psychiatrists was assigned from one to three diagnoses and told to complete the protocols so as to describe hypothetical subjects with those diagnoses. The psychiatrists were told that their protocols would be diagnosed by another psychiatrist as well as by DIAGNO and that they could think of the study as "Man Versus the Computer" in that the

computer's ability to recognize their hypothetical case would be compared with that of their fellow clinicians. Therefore, their protocols should contain sufficient information to enable the clinician to recognize the diagnosis but not be so stereotyped as to be easily recognized by both man and the computer. They were further challenged to attempt to "fake out" the computer but were warned that if their case was too atypical, the clinician would also fail to correctly diagnose their case.

The 30 raters completed 54 such hypothetical cases (two for each diagnosis). Each completed protocol was given to another psychiatrist from among the 30 raters. The second psychiatrist then made a diagnosis

**Table 5.—Distribution of 200 Real Cases by Major Diagnostic Groups for Ten Clinicians and DIAGNO**

Diagnosis	Average for Clinicians	DIAGNO
Organic brain syndrome	7.3	2.0
Psychotic depressive reaction	16.5	5.5
Schizophrenia	38.0	27.0
Neurosis	21.0	28.0
Personality disorder	4.8	0.0
Sociopathic disorder	14.3	18.0
Nonspecific illness	1.8	11.5
Not ill	0.5	8.0
Other	0.8	0.0

**Table 6.—Chance Corrected Diagnostic Agreement (Weighted Kappa) for the 50 PSS Protocols With a Diagnosis Made by the Interviewer**

	Packet*					Median
	A	B	C	D	E	
Interviewer—Clinician 1	0.28	0.16	0.15	0.44	0.33	0.28
Interviewer—Clinician 2	0.21	0.36	0.24	0.49	0.19	0.24
Interviewer—DIAGNO	0.45	0.25	0.19	0.30	0.25	0.25

The 50 cases contained 20 former psychiatric outpatients, 20 community cross-section subjects, and 10 prisoners.

\* Each packet contained 10 cases.

**Table 7.—Distribution of Diagnoses Given to the 20 Nonpatients by Interviewers, Clinicians, and DIAGNO**

	Interviewers %	Clinicians* %	DIAGNO %
Psychotic depressive reaction	0	10	0
Schizophrenia	0	7.5	0
Personality disorder	30	15	0
Alcoholism or drug addiction	10	0	0
Neurosis	25	55	20
Psychophysiological reaction	10	0	0
Transient situational reaction	5	0	0
Nonspecific illness with mild symptomatology	0†	10	20
Not ill	20	2.5	60
Total	100	100	100

\* Based on 40 diagnoses from two clinicians for each case.

† This category was not available to the interviewers.

using one of the 27 categories. Each psychiatrist completed and diagnosed from one to three cases.

The originally assigned hypothetical diagnosis, the second clinician's diagnosis, and DIAGNO's diagnosis were compared with each other. The chance corrected agreement, weighted kappa, was 0.54 between the hypothetical and clinical diagnoses, 0.59 between the hypothetical and DIAGNO diagnoses, and 0.52 between the clinical and DIAGNO diagnoses. Thus, DIAGNO recognized the hypothetical cases as well as did the clinicians. Furthermore, DIAGNO agreed with both the psychiatrists' hypothetical and clinical diagnoses as well as they agreed with each other.

This shows that when hypothetical cases are described so that stereotypes are avoided, yet they contain enough information to make a diagnosis, DIAGNO correctly identifies the cases as well as clinicians. Since the average disagreement level for the cases involving the three attending psychiatrists was no lower than the average level for the psychiatric residents, it is unlikely that these results would be different if more experienced clinicians had been used. Undoubtedly the agreement would have been greater had the hypothetical cases been constructed so as to describe the "typical" case.

The purpose of the second study was to see if DIAGNO diagnoses agree with a psychiatrist's diagnoses as well as two psychiatrists agree with each other when DIAGNO and the two psychiatrists are all limited to information from PSS protocols. The PSS protocols of 200 actual interviews of a heterogeneous group of subjects were used. The subjects consisted of 125 psychiatric inpatients (75 admissions to three St. Louis psychiatric hospitals or services, and 50 admissions to the New York State Psychiatric Institute), 25 men interviewed at the Municipal Shelter in the Bowery section of New York City, 10 prisoners, 20 former psychiatric outpatients, and 20 nonpatients in the Washington Heights section of New York City. (The data on the prisoners, outpatients, and nonpatients have been provided by Bruce Dohrenwend, PhD, and were gathered by psychiatrists who were trained to administer the PSS. It represents only an initial portion of data being collected by him

in his studies. The interpretation of the data is made by the authors alone.) These cases were divided into five packets of 40 cases each, such that the packets contained the same number of cases from each of the above samples.

Each packet was given to one of five pairs of experienced board certified or eligible psychiatrists who were engaged in research or teaching that involved the use of the APA nomenclature. They were told that some of the cases would be prisoners, nonpatients living in the community, etc, but not what proportion of cases were from the various samples.

Each of the pairs of psychiatrists independently examined each protocol and made a diagnosis. They were not limited to the 27 DIAGNO diagnoses but were free to use any standard APA diagnosis and qualifying phrase as well as not ill and nonspecific illness with mild symptomatology.

The weighted kappas for agreement between the two clinicians and between each clinician and DIAGNO for the five packets are shown in Table 4. The agreement was lower than that found by Sandifer when clinicians diagnosed patients whom they actually saw and on whom they had historical information. However, since it was only slightly lower, it indicates that using the PSS protocols for making a diagnosis, while not the same as the usual clinical diagnostic process, is not so different as to be irrelevant.

For four of the five packets, the clinicians agreed somewhat better with each other than with DIAGNO. This shows that for real cases, as contrasted with hypothetical cases, DIAGNO diagnoses do not agree with psychiatrists' diagnoses as well as they agree with each other. DIAGNO agreed best with the pair of psychiatrists who had the greatest agreement between themselves (packet A). Similarly, DIAGNO had the lowest agreement with the pair of psychiatrists who had the lowest agreement between themselves (packet E).

The cases in which the clinicians agreed with each other better than with DIAGNO were examined. In about a third of the cases the clinicians were apparently able to agree on a diagnosis by utilizing one or more key items which DIAGNO, being limited to

Table 8.—Diagnoses for 57 Patients by DIAGNO and an Attending Psychiatrist

DIAGNO	Attending Psychiatrist	Level of Disagreement
Paranoid schizophrenia, behavior reaction	Schizo-affective schizophrenia, behavior reaction	2
Depressive neurosis	Depressive neurosis	0
Chronic undifferentiated schizophrenia, behavior reaction	Chronic brain syndrome due to alcohol, psychotic reaction	7
Paranoid schizophrenia	Acute undifferentiated schizophrenia	2
Sociopathic personality, alcoholism, neurotic reaction	Depressive neurosis, behavior reaction	5
Depressive neurosis	Paranoid state	7
Paranoid schizophrenia	Paranoid schizophrenia	0
Sociopathic personality, alcoholism	Depressive neurosis	5
Depressive neurosis	Depressive neurosis	0
Depressive neurosis	Passive aggressive personality	5
Anxiety neurosis	Chronic undifferentiated schizophrenia	9
Chronic undifferentiated schizophrenia, neurotic reaction	Depressive neurosis	7
Sociopathic personality, alcoholism, neurotic reaction	Acute brain syndrome due to alcohol, psychotic reaction	3
Sociopathic personality, alcoholism, neurotic reaction	Chronic undifferentiated schizophrenia, neurotic reaction	7
Paranoid schizophrenia	Paranoid schizophrenia	0
Chronic undifferentiated schizophrenia	Catatonic schizophrenia	2
Sociopathic personality, alcoholism	Psychotic depressive reaction	7
Depressive neurosis	Depressive neurosis	0
Paranoid schizophrenia, behavior reaction	Schizo-affective schizophrenia	2
Paranoid schizophrenia, neurotic reaction	Paranoid schizophrenia, mental deficiency	1
Chronic undifferentiated schizophrenia, behavior reaction	Sociopathic personality, drug addiction with psychotic reaction	7
Sociopathic personality, alcoholism	Adolescent adjustment reaction, neurotic reaction	5
Sociopathic personality, alcoholism	Sociopathic personality, alcoholism	0
Paranoid schizophrenia, behavior reaction	Chronic undifferentiated schizophrenia	2
Paranoid schizophrenia, behavior reaction	Chronic undifferentiated schizophrenia	2
Anxiety neurosis	Chronic undifferentiated schizophrenia, behavior reaction	9
Sociopathic personality, antisocial, neurotic reaction	Chronic undifferentiated schizophrenia, behavior reaction	7
Sociopathic personality, alcoholism, neurotic reaction	Paranoid schizophrenia	7
Depressive neurosis	Chronic undifferentiated schizophrenia	9
Nonspecific illness with mild symptomatology	Depressive neurosis	5
Paranoid schizophrenia, neurotic reaction	Paranoid schizophrenia	1
Paranoid schizophrenia	Paranoid state, neurotic reaction	3
Dissociative neurosis	Acute undifferentiated schizophrenia	7
Schizo-affective schizophrenia, behavior reaction	Chronic undifferentiated schizophrenia, behavior reaction	2
Chronic undifferentiated schizophrenia, neurotic reaction	Chronic undifferentiated schizophrenia, neurotic reaction	0
Nonspecific illness with mild symptomatology	Chronic undifferentiated schizophrenia	7
Paranoid schizophrenia, neurotic reaction	Acute undifferentiated schizophrenia	2
Paranoid schizophrenia, behavior reaction	Paranoid schizophrenia	1



Table 8.—Diagnoses for 57 Patients by DIAGNO and an Attending Psychiatrist (Continued)

DIAGNO	Attending Psychiatrist	Level of Disagreement
Chronic brain syndrome, cerebral arteriosclerosis	Involuntal psychosis	5
Depressive neurosis	Acute brain syndrome, trauma	9
Depressive neurosis	Psychotic depressive reaction	3
Sociopathic personality, alcoholism	Sociopathic personality, alcoholism, neurotic reaction	1
Chronic brain syndrome, cerebral arteriosclerosis, psychotic reaction	Paranoid schizophrenia, neurotic reaction	7
Depressive neurosis	Adolescent adjustment reaction, behavior reaction	3
Psychotic depressive reaction	Manic depressive depressed	3
Sociopathic personality, drug addiction, neurotic reaction	Sociopathic personality, drug addiction, psychotic reaction	1
Psychotic depressive reaction, behavior reaction	Sociopathic personality, drug addiction, neurotic reaction	7
Sociopathic personality, alcoholism, neurotic reaction	Chronic undifferentiated schizophrenia, neurotic reaction	7
Nonspecific illness with mild symptomatology	Chronic brain syndrome, unknown cause	9
Nonspecific illness with mild symptomatology	Adult situational reaction, behavior reaction	5
Chronic undifferentiated schizophrenia	Paranoid schizophrenia	2
Chronic undifferentiated schizophrenia, behavior reaction	Paranoid schizophrenia	2
Paranoid schizophrenia, neurotic reaction	Acute undifferentiated schizophrenia	2
Paranoid schizophrenia	Paranoid schizophrenia	0
Psychotic depressive reaction	Paranoid schizophrenia	7
Chronic undifferentiated schizophrenia	Chronic undifferentiated schizophrenia	0
Paranoid schizophrenia	Paranoid schizophrenia	0

scale scores, ignored. For example, several patients were diagnosed by both clinicians as psychotic because of the presence of one or two very suggestive items such as "Speaks of his problem with no signs of emotion." There was also a tendency for the clinicians to combine behaviors from several scales where each scale failed to reach a DIAGNO cut-off level, to ignore items which were inconsistent with the overall clinical picture yet which were used by DIAGNO to make a different diagnosis, to take age into account as suggesting organicity, and to diagnose brain syndromes due to alcohol when there was no disturbance in memory or orientation.

The distribution of the clinicians' diagnoses by major diagnostic group differed from that of DIAGNO (Table 5). The clinicians more often diagnosed organic brain syndrome, psychotic depressive reaction, and schizophrenia, whereas DIAGNO ap-

pears to be more conservative in that it more commonly diagnosed neuroses, nonspecific illness, or not ill.

The next study evaluated DIAGNO by comparing both DIAGNO and the clinicians against a third criterion: a psychiatrist who made a diagnosis after he completed a PSS interview. There were 50 protocols of subjects with an interviewer's diagnosis: the 20 former psychiatric outpatients, the 10 prisoners, and the 20 nonpatients. The two clinicians' and DIAGNO's diagnoses were compared with those of the interviewer (Table 6). Although the agreement was generally low, the agreement between the interviewers' and DIAGNO's diagnoses was no lower than that between either clinician and the interviewer. Thus, while the clinicians agreed better with each other than with DIAGNO, they agreed no better than DIAGNO with the psychiatrist who conducted the interview.

The distribution of diagnoses given to the 20 nonpatients (Table 7) showed that while 17.5% of the clinicians' diagnoses were either schizophrenia or psychotic depressive reaction, neither the interviewer nor DIAGNO made these diagnoses for any of these subjects. Furthermore, whereas only 2.5% of the clinicians' diagnoses were not ill, the interviewer and DIAGNO made this diagnosis for 20% and 60% of the subjects, respectively. There are at least two alternative interpretations of these results: either DIAGNO fails to detect significant psychopathology, or the clinicians share a common bias towards labeling a person with a few symptoms as psychiatrically ill. However, none of the 125 psychiatric inpatients, and only three of the 20 former outpatients, were diagnosed by DIAGNO as not ill. We therefore view the latter interpretation as the more likely.

Finally, DIAGNO diagnoses from PSS interviews of 57 consecutive admissions to the Washington Heights Community Service of the New York State Psychiatric Institute were compared with the diagnoses made by the attending psychiatrists after an admission conference (Table 8). The weighted kappa for agreement was 0.30.

For 13 of the 20 cases with the greatest disagreement, the attending psychiatrist gave the more serious diagnosis. In ten of these instances the attending psychiatrist made a diagnosis of chronic undifferentiated or paranoid schizophrenia, presumably on the basis of historical information or more subtle signs not detected by DIAGNO. The agreement was not as high as would be desired or as high as usually obtained between clinicians present at the same interview. Nevertheless, it demonstrates the ability of DIAGNO to make use of data on current status, collected in a structured research interview by nonpsychiatrists, to approximate the diagnoses arrived at by psychiatrists in the usual clinical procedure.

#### Comments

How do the studies and results reported here compare with previous attempts at computerized psychiatric diagnosis? Overall and his associates<sup>3-5</sup> have utilized four different statistical models in developing computer programs for psychiatric diag-

nosis. These programs differ in several significant respects from DIAGNO. First of all, they were developed on the basis of clinicians' descriptions of "typical" cases of particular diagnoses rather than on data from real cases. Furthermore, these programs are limited to only 13 functional psychotic diagnoses. Smith's work<sup>6</sup> is a modification and extension of the pattern probability procedure, one of the four statistical models used by Overall. Like Overall, his program was also developed on the basis of clinicians' descriptions of "typical" cases. Smith's program considers 38 different diagnoses, some of which are not standard APA diagnoses (eg, borderline schizophrenia), and other diagnoses which are not part of the output of DIAGNO (eg, six psychophysiological reactions). Neither Overall's nor Smith's programs have provision for the diagnosis of not ill or nonspecific illness. As a consequence, their programs assign a psychiatric diagnosis to all subjects.

Part of the validation of Overall's and Smith's programs, as well as of DIAGNO, was to test the ability of the program to identify hypothetical cases. In this regard, both Smith's and Overall's programs were somewhat more successful than DIAGNO. However, it should be noted that their hypothetical cases were "typical" cases, whereas the clinicians who constructed the hypothetical cases used to validate DIAGNO were told to avoid "typical" cases, and in most instances did so. A more important difference, however, is that the hypothetical cases used to test Overall's and Smith's programs were the very same cases that were used to develop their programs. In contrast, DIAGNO was developed on real cases and the hypothetical cases were a cross validation procedure.

All three programs have been tested with real cases. However, in the studies of both Overall and Smith the clinical and the computer diagnoses were not made completely independently. In Smith's study of real patients, he made both a clinical diagnosis and filled out the symptom ratings that were the input for his program for 30 hospitalized psychiatric patients (oral communication, Dec 22, 1965.). His report of 87% agreement (not chance corrected) between his diagnosis and the computer diagnosis is dif-

difficult to interpret because of his familiarity with the computer program.

The agreement found between DIAGNO's diagnoses and clinical diagnoses for real patients is greater than that reported by Overall<sup>5</sup> for clinical diagnoses of a group of real patients and the diagnoses from one of his computer programs. He compared the clinical diagnoses for 314 patients who had one of his 13 programmed diagnoses with the diagnoses derived from a computer program based on one of the profile similarity classification models. Although the clinicians were not familiar with the computer program, approximately half of the cases diagnosed by the same psychiatrist who completed the behavior ratings used by the program (oral communication, April 8, 1967). Nevertheless, the unweighted kappa for chance corrected agreement between the computer and the clinical diagnoses was only 0.06. (This was computed by us from Overall's data. Unweighted kappa has the same meaning as weighted kappa except that all disagreements are equally weighted. A comparison of the weighted kappa for the two studies is not appropriate. In Overall's study both the computer and the clinician were limited to 13 diagnoses which for the most part have low levels of diagnostic disagreement with each other, whereas the diagnoses used in the DIAGNO studies make it possible to have the full range of levels of diagnostic disagreement.) The unweighted kappa for chance corrected agreement between DIAGNO and the clinical diagnoses for the 57 cases is 0.22.

The modest results reported here could undoubtedly be improved if DIAGNO had, as input, data relevant to history and the development of the present illness. (Work has begun on such a program, DIAGNO II.) This would enable it to better diagnose cases in which there are features of more than one illness, for example, acute alcoholism in a chronic schizophrenic. The absence of historical information requires DIAGNO to make certain decisions on the basis of rather arbitrary criteria (eg, involuntional reactions are separated from psychotic depressive reactions solely on the basis of age and sex), and made it impossible to diagnose many of the personality disorders.

We have demonstrated that a computer program for psychiatric diagnosis, based on

a logical decision tree approach similar to the differential diagnostic model used in clinical medicine approximates the clinical diagnostic procedure. Possible uses of DIAGNO or other programs using this model include selecting subjects for study or assignment to different treatment regimens or describing samples used in studies.

A study by Kriegel<sup>10</sup> on reaction time illustrates its use as a selection procedure. There were 19 patients with a hospital diagnosis of schizophrenia and 22 paid "normals" who were interviewed with the PSS. The 11 patients who also had a DIAGNO diagnosis of schizophrenia had a smaller variability in reaction time, that is, were more homogeneous, than the entire patient group. Similarly, the 11 subjects from the normal group who also had a DIAGNO diagnosis of not ill were more homogeneous than the entire normal group. The subjects in the normal group who had a DIAGNO diagnosis indicative of some psychiatric disturbance behaved more like the patient group than the 11 subjects whose DIAGNO diagnosis was not ill.

Another use of this approach would be in epidemiological studies examining the prevalence of psychiatric disorders in the community. Common problems in such studies are the cost of using psychiatrists as field interviewers and the lack of uniformity in the criteria used to define a case. DIAGNO is based on input from a standardized interview schedule which can be effectively administered by nonpsychiatrists. The criteria for assigning a psychiatric diagnosis would be uniform across all subjects.

The output of DIAGNO is a standard APA diagnosis. Several investigators<sup>11-13</sup> have suggested that nonstandard typologies may be far more useful for selecting optimal treatment or predicting treatment response. Programming for such typologies does not have to be limited to probabilistic or other statistical models. For example, Klein<sup>13</sup> reports a computer program which utilizes a "successive sieving" model for classifying patients to optimize choice of drug treatment. In a similar fashion, the differential diagnostic model used in DIAGNO could also be the basis of programs which classify patients according to nonstandard typologies.

## Summary

In recent years there has been increased interest in using computers to arrive at a clinical diagnosis. Although several attempts at developing computer programs for classifying patients according to the standard APA psychiatric nomenclature have been made, they have all relied on various statistical models. This paper describes a computer program for psychiatric diagnosis, DIAGNO, which is based on a logical decision tree model similar to the differential diagnostic procedure employed in clinical medicine.

The input data for DIAGNO are limited to the information on current psychopathology, contained in scale scores of the Psychiatric Status Schedule, as well as age, sex, and number of previous psychiatric hospitalizations. The output is one of 25 standard APA diagnoses and qualifying phrases as well as two unofficial diagnoses: not ill and nonspecific illness with mild symptomatology.

Reliability and validity studies are reported indicating that even at this early stage in the development of computerized psychiatric diagnosis, the method described here is of sufficient validity to be of research use.

This investigation was supported in part by Public Health Service grant MH 08534 from the National Institute of Mental Health.

A portion of the data was provided by Drs. Ralph Biddy (St. Louis patients), and Bruce Dohrenwend (community subjects).

Copies of the Psychiatric Status Schedule, a list of the items comprising the scales, program listings, and DIAGNO in narrative form are available upon request.

## References

1. Beck, A.T., et al: Reliability of Psychiatric Diagnosis: II. A Study of Consistency of Clinical Judgment and Rating, *Amer J Psychiat* 119:351-357, 1962.
2. Sandifer, M.G.; Pettus, C.; and Quade, D.: A Study of Psychiatric Diagnosis, *J Nerv Ment Dis* 139:350-356, 1964.
3. Overall, J.E.: A Configural Analysis of Psychiatric Diagnosis, *Behav Sci* 8:211-219, 1963.
4. Overall, J.E., and Gorham, D.R.: A Pattern Probability Model for the Classification of Psychiatric Patients, *Behav Sci* 8:108-116, 1963.
5. Overall, J.E., and Hollister, L.E.: Computer Procedures for Psychiatric Classification, *JAMA* 187:583-588, 1964.
6. Smith, W.G.: A Model for Psychiatric Diagnosis, *Arch Gen Psychiat* 14:521-529, 1966.
7. *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Washington, DC, 1952.
8. Cohen, J.: A Coefficient of Agreement for Nominal Scales, *Educ Psychol Meas* 20:37-46, 1960.
9. Spitzer, R.L., et al: Quantification of Agreement in Psychiatric Diagnosis: A New Approach, *Arch Gen Psychiat* 17:83-87, 1967.
10. Kriegel, J.: *Reaction Time in Schizophrenics and Normals as a Function of Stimulus, Uncertainty, Guessing, and Modality Shift*, dissertation, Columbia University, New York, 1967.
11. Goldberg, S.C., et al: Prediction of Improvement in Schizophrenia Under Four Phenothiazines, *Arch Gen Psychiat* 16:107-117, 1967.
12. Katz, M.M.; Lowery, H.A.; and Cole, J.O.: "Behavior Patterns of Schizophrenics in the Community," in Lorr, M. (ed.): *Explorations in Typing Psychotics*, Oxford, England: Pergamon Press, 1966, pp 209-230.
13. Klein, D.F.; Honigfeld, G.; and Feldman, S.: "Prediction of Drug Effect by a Successive Sieving Diagnostic Technique," in Wittenborn and May (eds.): *Prediction of Response to Pharmacotherapy*, Springfield, Ill: Charles C Thomas, Publishers, to be published.