Eur Arch Psychiatr Neurol Sci (1989) 239:127–132

European Archives of Psychiatry and Neurological Sciences

On the Way to Expert Systems

Comparing DSM-III Computer Diagnoses with CATEGO (ICD) Diagnoses in Depressive and Schizophrenic Patients*

K. Maurer, H. Biehl, C. Kühner, and W. Löffler**

Zentralinstitut für Seelische Gesundheit, Postfach 122120, D-6800 Mannheim 1, Federal Republic of Germany

Summary. Operationalized diagnostics deal with the standardized assessment of psychiatric symptoms as well as diagnostic criteria. As a diagnostic system based on criteria, the DSM-III was chosen to identify operationalized diagnoses based on the Present State Examination (PSE-9) and some additional DSM-III specific items. By relating PSE symptoms to the diagnostic criteria of DSM-III, an easily applicable expert system leading to DSM-III diagnoses was developed. In two samples of 30 schizophrenic and 51 depressive patients the DSM-III computer diagnoses are contrasted with the ICD-8 diagnoses of the PSE/CATEGO system. In defining a "case", only minimal differences between the two computer programs were found. In the sample of schizophrenics, CATEGO led to 114 (81%) diagnoses and the DSM-III program to 112 (79%) diagnoses; for the depressive patients 43% cases were identified by CATEGO and 45% by the DSM-III algorithm. Comparing the diagnosis of "acute schizophrenic disorders", both programs arrived at similar percentages. (CATEGO: 51%; DSM-III: 57%). However, CATEGO is limited to two different subtypes (295.2 and 295.3), whereas the DSM-III program covers the total range of possible schizophrenic subtypes. Furthermore, the DSM-III program identified residual subtypes of schizophrenia in 23% of the diagnostic decisions. In the short-term course of the schizophrenic patients, CATEGO identified 27%–43% with affective diagnoses with high stability per cross-section. Using the DSM-III algorithm affective diagnoses were rather rare (maximum of 17%), marking unstable changes from acute to residual states of the psychosis. In the cohort of affective patients the correspondence between both programs was quite good, especially for the patients with a diagnosis of depression, but in total the DSM-III program requires stricter criteria for affective – especially manic – disorders, whereas CATEGO needs a somewhat higher symptom level for anxiety syndromes to reach the diagnostic threshold.

Key words: Computer-assisted diagnosis – DSM-III algorithm – PSE/CATEGO system

Introduction

One consequence of the debate on the reliability of diagnoses (e.g. Kreitman et al. 1961; Shepherd et al. 1968; Jakubaschk and Werner 1974, 1976) has been the development of "operationalized diagnostics", thus making the diagnostic process more clear and replicable. Two different aspects of the diagnostic process must be taken into account: the assessment of psychiatric symptoms – what Wing called "defining the data base" (Wing 1984) – and the application of sets of rules to combine the identified symptoms to form a diagnostic statement. Both components contribute specific sources of error, i.e. observation vs information variance (Spitzer and Fleiss 1974). The former concerns the identification and evaluation of symptoms, the latter the "criterion variance" of the application of rules combining symptoms to reach a psychiatric diagnosis. These sources of error can be reduced: (1) by (semi-) standardized diagnostic interviews, e.g. the PSE (Wing et al. 1974), to assure that

* Modified version of a paper presented by K. Maurer at the 2nd Psychiatric Autumn Symposium of the ZISG at St. Martin, Palatinate, 15–17 September 1988

** Funded by the ZISG (Zentralinstitut für Seelische Gesundheit) and the BMFT (Bundesministerium für Forschung und Technologie) under the coordination of the GSF (Gesellschaft für Strahlenforschung)

Offprint requests to: K. Maurer

different diagnosticians consider and elicit the same range of symptoms, identify them with explicit rules and rate their intensity in an adequate, fairly operationalized and reliable manner; (2) by the application of obligatory diagnostic criteria, conventional and testable rules guiding the diagnostician as to which symptoms are required to reach a specific diagnosis (inclusion criteria) and which symptoms or alternative diagnoses preclude a diagnosis (exclusion criteria).

Ideally a standardized data base and operationalized diagnostic criteria, e.g. RDC (Spitzer et al. 1975), DSM-III (APA 1980), are the prerequisites for a diagnostic algorithm, which then can be realized as a computer program. The term "operationalized diagnostics" is often reserved for computer diagnostics.

Diagnostic Algorithms as Expert Systems

One discipline in computer science which deals with the "simulation" or adaptation of typical human cognitive processes such as decision-making, searching for optimal strategies etc. by a machine is designated "artificial intelligence" (AI) (Waltz 1982; Stede 1984; Hand 1985). One important application of AI is the creation of expert systems. These are computer programs that have expert knowledge of a special discipline in a "knowledge base" (Hayes-Roth et al. 1983; Kalb 1986). Rules are the most important part of the knowledge base; these are representations of the experts' knowledge that have to be combined with the data base, a set of specific data or the data of a single person, which the potential user has to contribute. For example, if the criteria of a diagnostic system are to be the knowledge base, then the symptoms of a certain proband may be regarded as the data base. The diagnostic expert system combines symptoms and rules and reaches a decision about the diagnosis of the proband. Diagnostic algorithms are evident applications of expert systems in medicine, especially in psychiatry.

The need for a DSM-III Algorithm Based on PSE data

Apart from the need of communicating with the DSM-III community, our main interests for developing a DSM-III computer program were based on the shortcomings of the existing CATEGO program, which were hard to reconcile or to explain:

1. In a cohort of 70 first-onset schizophrenics, CATEGO identified on the average about 20% affective diagnoses in seven cross-sections over 5 years, two-thirds of them with depression, one-third with mania (Biehl et al. 1985).

2. The diagnostic range of the CATEGO program did not include „residual schizophrenia", because this diagnosis is too narrowly defined as a syndrome and cannot be based on "present state items" alone (Wing et al. 1974; Biehl et al. 1988).

One possibility of defining residual schizophrenia by (re-) combination of PSE symptoms was to identify a "residual factor" by principal component analysis with high loadings of typical minus (or "negative") symptoms such as lost emotions, anergia, retardation, etc. We found a stable three-factor solution in two independent cohorts of schizophrenic patients with a delusional, a residual and an anxiety factor (Krüger et al. 1985, 1988; Biehl et al. 1988).

Another way of identifying "negative schizophrenia" was to develop an alternative computer program to the PSE/CATEGO system using the given PSE data base and precisely defined diagnostic criteria in order to achieve clear rules for symptom combinations. For this strategy the existing DSM-III (APA 1980) seemed suitable for an empirical test, because it consists of widely accepted operationalizations of diagnostic criteria but lacks any convention concerning the collection of data.

The Steps of Program Development

Initially we worked on a DSM-III algorithm which was a simple BASIC program for an 8-bit processor. Later it was expanded and transformed into FORTRAN-77 for a main frame computer. A program test with ten psychotic cases (Biehl et al. 1986) made the need for some program revisions obvious: the operationalizations of the diagnostic criteria had to be checked and formulated more precisely; the diagnostic spectrum had to be extended, e.g. for non-psychotic diagnoses; the data input from a data file was supplemented by data input by screen; also the program output was organized in a user-adequate form, containing a "verbalized" diagnostic statement; additionally, the parameters computed by the program, i.e. the diagnostic criteria and diagnoses, could be stored in a data file for further analyses.

After formal program testing the different diagnoses were validated by simulation data, starting with "prototype diagnoses" (Cantor et al. 1980; de Jong and van den Brink 1986). After these program tests, the algorithm was used to analyse two empirical data sets to compare the DSM-III diagnoses with conventional CATEGO results, leading to tentative ICD diagnoses.

The Structure of the DSM-III Program

The program consists of four essential parts: In the first part, it runs through an input or data definition

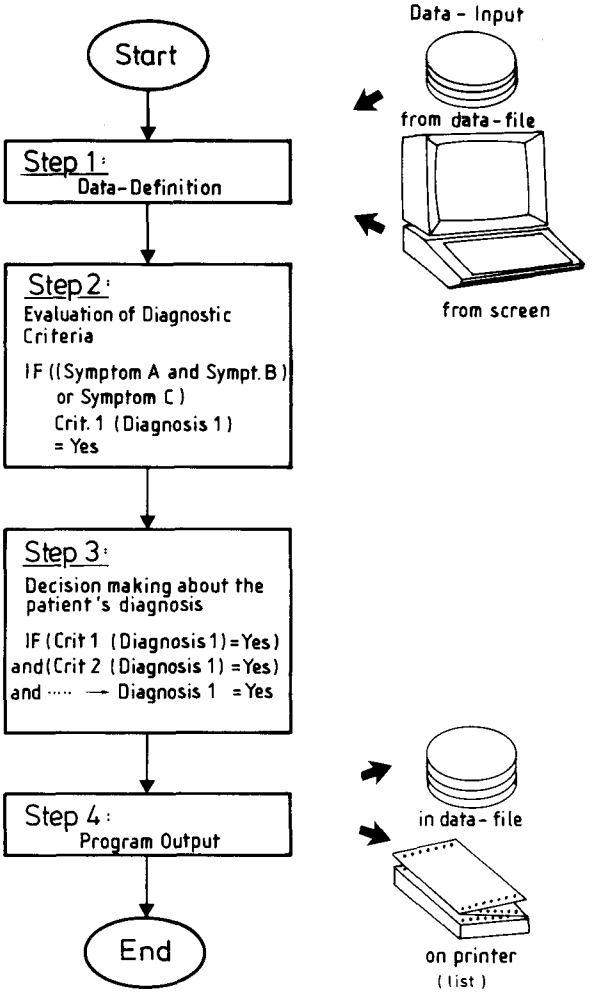

Fig. 1. The structure of the DSM-III algorithm

routine, where data input is possible from a data file or directly on the screen. The data input consists of 140 PSE symptoms and 22 additional DSM-III-specific items. The central part of the program checks the diagnostic criteria, including the diagnoses of schizophrenic disorders with the diagnostic subtypes: disorganized, catatonic, paranoid, undifferentiated and residual, the paranoid disorders, and psychotic disorders not classified elsewhere. This central part also covers affective disorders, namely major affective disorders with coding of the fifth digit (with psychotic features, in remission, with melancholia, without melancholia and other specific affetive disorders), anxiety disorders including phobias (agoraphobia, with or without panic attacks; social phobia; simple phobia) and anxiety syndromes (panic syndrome, generalized anxiety syndrome, obsessive-compulsive syndrome). Part 3 of the program is called the "decision module" and brings all the possible diagnoses (e.g. those with a set of fulfilled criteria into a hierarchical order implicit in the DSM-III, suggesting the diagnosis in the top position (e.g. schizophrenia or major depressive disorder over phobia or panic disorder). Part 4 controls the program output and supplies the user with a diagnostic statement (e.g. "patient no. 20 has a paranoid disorder, 297), either on the screen or in a printout. Additionally, the program creates a data file output, containing the input raw data, the computed DSM-III criteria, some sum-scores for specific psychopathologies needed to evaluate certain diagnostic criteria, and also the DSM-III diagnoses for all input cases. This data file can be used for further statistical analyses or as the starting point for "polydiagnostics" (e.g. Berner and Katschnig 1983; Philipp and Maier 1987).

Results

Cross-sectional Diagnoses of 51 Depressive Patients

For 51 depressive patients 4 weeks after discharge[1] we compared the CATEGO-ICD diagnoses (Wing et al. 1974) and the DSM-III computer diagnoses.

In identifying a "case" only minimal discrepancies were found between the two programs. In 57% of the patients the symptomatology available was not adequate to produce a diagnostic statement using the CATEGO program (Index of Definition less than 5). For the DSM-III algorithm the corresponding percentage was 55%. Comparing the depressive diagnoses we also found a high rate of corresponding diagnostic judgements: CATEGO suggested 31.3% psychotic (296.2) or neurotic depressions (300.4), and the DSM-III program identified 33.3% (major) depressions. One case was labelled as an atypical psychotic disorder in both diagnostic systems.

Discrepancies were found for three patients with CATEGO mania. This diagnosis never occurred using

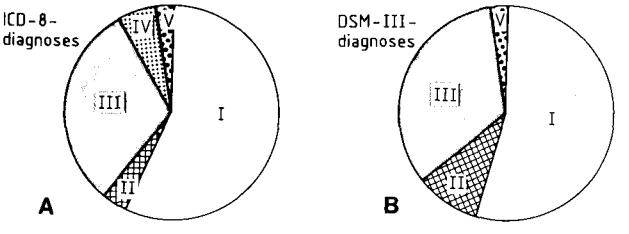

Fig. 2. Frequencies of operationalized ICD-8-CATEGO and DSM-III computer diagnoses of 51 depressive patients (4 weeks after hospital discharge). **A** I = no diagnosis; II = anxiety disorders; III = depression; IV = mania; V = atypical psychosis. **B** I = no diagnosis; II = anxiety disorders; III = major depression; V = atypical psychosis

[1] Funded by the Deutsche Forschungsgemeinschaft. The data were made available by Dr. H. O. F. Veiel

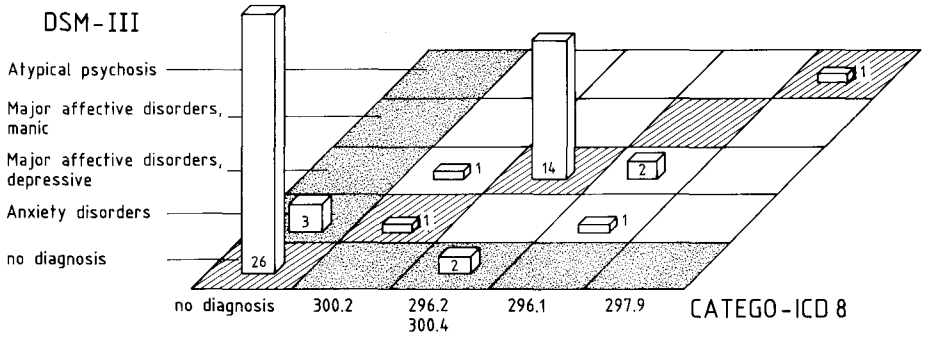

Fig. 3. Correspondence of operationalized CATEGO-ICD-8 and DSM-III diagnoses of 51 depressive patients (4 weeks after hospital discharge). 82.4% ▨ Correspondence (51% no cases, 31.4% cases); 9.8% ■ only one of the two programs results in a diagnosis; 7.8% □ no diagnostic correspondence

the DSM-III program, although it is operationalized in the algorithm. Further differences were apparent in the anxiety disorders: while CATEGO identified only 4% of the patients with "anxiety neurosis", the DSM-III program found 10%.

Bivariate distribution of the diagnoses of the two programs resulted in 82% concordance, the main proportion being "no diagnosis". Of the total cohort 27% were identified as depressive by both programs; also, one patient with acute psychotic symptoms and one patient with anxiety symptoms received corresponding diagnoses.

There were 5 cases diagnosed in only one of the systems. Two patients had a depression by CATEGO that did not fulfil the diagnostic criteria for a DSM-III diagnosis. Three patients had a DSM-III anxiety disorder, where the Index of Definition was too low to produce a tentative ICD diagnosis by CATEGO.

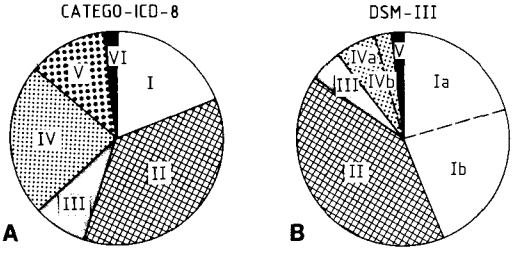

Fig. 4. Frequencies of operationalized ICD-8-CATEGO and DSM-III computer diagnoses of 141 PSE interviews with 30 schizophrenic patients. **A** I = no diagnosis; II = schizophrenia; III = paranoid syndromes; IV = endogenous/neurotic depression; V = mania; VI = anxiety neurosis. **B** Ia = no diagnosis; Ib = residual syndrome; II = schizophrenic/schizophreniform disorders; III = paranoid disorders; IVa = major depression, with psychotic features; IVb = major depression; V = anxiety disorders

Short term Course of 30 Schizophrenic Patients

The "short-term" course of 30 acute schizophrenic patients, 15 of them first-onset patients and 15 of them with a more chronic course, was followed over at least 4 months, beginning with the admission to hospital until at least 6 weeks after discharge from hospital. On admission the patients' mental state was assessed by the PSE; this was repeated every 4 weeks, and at least one PSE was performed after discharge at each patient's home. The PSE/CATEGO results in the short-term course of the schizophrenic cohort will now be compared with the results of the DSM-III algorithm.

In terms of "case" identification, the two programs showed high agreement. Over all cross-sections, CATEGO identified 19% as "no diagnosis", the DSM programm 21%. However, comparing the diagnoses of acute episodes, we found quite typical differences between the two approaches.

Concerning schizophrenia, CATEGO only differentiated between 295.2, 295.3, and 297.9, whereas the DSM-III algorithm grouped the patients into five schizophrenic subtypes, paranoid psychoses and further diagnoses with acute symptoms not fulfilling the DSM-III criteria for schizophrenia (i.e. schizophreniform disorder, acute paranoid reaction, depressive or manic disorders with psychotic features). One problem seems to be the tendency towards a cross-sectional "affective" diagnosis of the CATEGO program. Thus we found a diagnosis of mania (296.1) in 17 cases, and even a diagnosis of depression (296.2) in 30 cases. In our opinion these clearly were patients either in an agitated acute or in a residual state of schizophrenia at the given cross-section.

Looking at these patients with "affective" diagnoses in the short-term course (Fig. 5), we found a relatively high proportion even on admission (first cross-section) using the CATEGO program, and 3 months later the diagnosis of an affective disorder was made in 13 of 30 patients. The stability of this diagnosis was quite high from one measurement to the next. The proportion of acute or "nuclear" schizophrenia diminished rapidly over time, and the proportions of patients without a CATEGO diagnosis increased, pointing to the important issue of "timing" a diagnostic interview in research at an optimal timepoint early in an episode of psychosis. This is also illustrated by the change in DSM-III diagnoses in the short-time course. For the DSM-III algorithm, how-

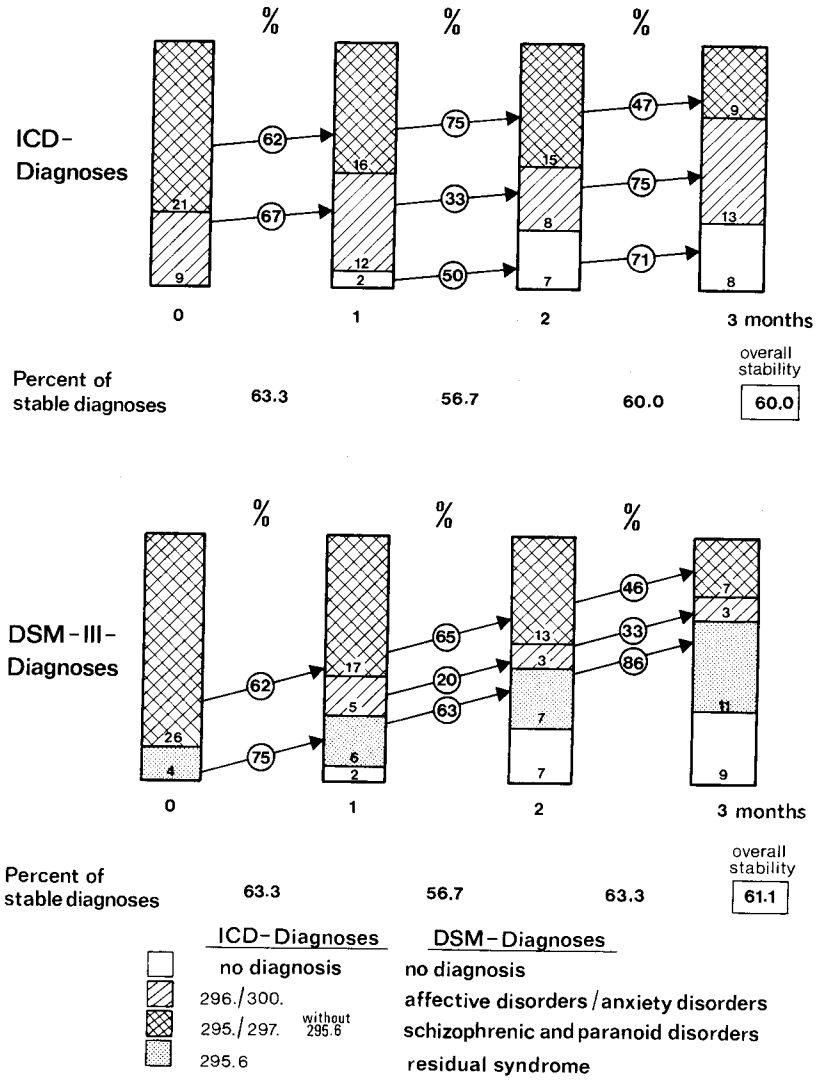

Fig. 5 Short-term course of patientes ($n = 30$) with acute psychotic symptoms (over 3 months after hospitalization). Frequency and stability of operationalized diagnoses

ever, the proportion of acute diagnoses at the starting point was higher than with CATEGO, and the proportion of patients without specific diagnoses was also higher than for CATEGO. However, the DSM-III program has the important advantage of differentiating between "no diagnosis" and "residual subtype of a schizophrenic disorder" after the remission of acute symptoms. Thus, the problem of misclassification of schizophrenia as an affective disorder seems to be resolved in our DSM program. Here we also found some affective disorders, but their proportions were much lower than in the CATEGO program, and the lower stability between two measurements indicated acute "schizo-affective" states within an extended episode of psychosis as intermediate steps on the way to residual phases. In the latest follow-up we then found both affective disorders (depression, aphanisis) and some psychotic features.

Conclusions

It was not the intention of this paper to demonstrate any superiority of either DSM-III or ICD, or of this new DSM-III algorithm or the already existing and widely used CATEGO program. We feel, however, that the introduction of the DSM-III diagnostic algorithm deserves testing and debating in direct comparison with the well-proven and widely accepted CATEGO program. Some of the shortcomings we were confronted with as users of the established PSE/CATEGO system could be overcome by this approach to a DSM-III algorithm, i.e. by the better separation of "no diagnosis" and cases "in remission", and also the problems of "affectiveness" – the proneness of CATEGO towards affective diagnoses and, finally, a more practical and plausible identification of "residual" schizophrenics on the basis of clinical experience.

Another advantage of the DSM-III seems to be the operationalization of diagnostic criteria. It also takes into account information about the course of a disease which is often needed to reach a clinically meaningful longitudinal diagnostic statement.

The World Health Organization is preparing the 10th revision of the ICD (WHO 1988; Olbrich and Maurer 1988) also with research diagnostic criteria and better-defined guidelines for its clinical use (Sartorius et al. 1988). In a parallel effort, a new generation of assessment instruments will be developed, called the "SCAN System" (Schedules for Clinical Assessment in Neuropsychiatry) (Wing et al. 1989) with the PSE-10 as the core element. For operationalized diagnoses, the new CATEGO-V program will be conceived as a "polydiagnostic algorithm" that will allow ICD-9 or ICD-10 diagnoses as well as compatibility with DSM-III R diagnoses (APA 1987) and possibly also with diagnoses of specific national diagnostic systems. The PSE-10 contains particular modules with items to evaluate specific diagnostic criteria, similar to our additional DSM-III adapted items. The time frame of the "classical" present episode – 1 month before the interview – can also be modified by different time spans such as present episode, typical episode, first episode or whole life-time. This indicates that the "fathers" of the PSE have also recognized the need to update the PSE/CATEGO system in the light of developments in the field of psychiatric diagnostics in the last decade.

References

American Psychiatric Association (APA) (1980) Diagnostic and statistical manual of mental disorders (3rd edn, DSM-III). Washington, DC

American Psychiatric Assosciation (APA) (1987) Diagnostic and statistical manual of mental disorders (3rd edn revised, DSM-IIIR). Washinton, DC

Berner P, Katschnig H (1983) Principles of "multiaxial" classification in psychiatry as a basis of modern methodology. In: Helgason T (ed) Methodology in evaluation of psychiatric treatment. Proceedings of a workshop held in Vienna, 10–13 June 1981. Cambridge University Press, Cambridge

Biehl H, Maurer K, Jung E (1985) Entwicklung von Behinderungen bei Schizophrenen – Erfassung, Differenzierung und Vorhersage. Ergebnisbericht des BMFT-Projekts 01 ZX 072/6

Biehl H, Maurer K, Jung E, Krumm B, Schubart C (1986) Neuere Aspekte der Behinderungsforschung bei Schizophrenen – Praktische Anwendung, Versorgung und Klassifikation. Ergebnisbericht zum GSF/BMFT Projekt 0701608 6

Biehl H, Maurer K, Jung E, Krüger G, Schubart C (1988) Reported symptoms in schizophrenic patients within five years of the onset of illness. In: Dencker SJ, Kulhanek F (eds) Treatment resistance of schizophrenia. Vieweg, Braunschweig/Wiesbaden

Cantor N, Smith EE, de Sales French R, Mezzich J (1980) Psychiatric diagnosis as prototype categorization. J Abnorm Psychol 89:183–193

De Jong A, van den Brink W (1986) Developments in psychiatric classification; introduction of a prototype core classification. Tijdschr Psychiatr 28:426–442

Hand DJ (1985) Artificial intelligence and psychiatry. Cambridge University Press, Cambridge

Hayes-Roth F, Waterman DA, Lenat DB (1983) Building expert systems. Addison-Wesley, Reading, Mass.

Jakubaschk J, Werner J (1974) Die Zuverlässigkeit psychiatrischer Diagnosen bei einer Wiederholungsuntersuchung. Nervenarzt 45:305–311

Jakubaschk J, Werner J (1976) Konzepte psychiatrischer Diagnosen – eine Ergänzungsuntersuchung zur Reliabilität psychiatrischer Diagnosen. Arch Psychiatr Nervenkr 223:45–58

Kalb R (1986) "Expertensysteme" in der Psychiatrie. Psychiatr Neurol Psychother Klin Prax 12:537–544

Kreitman N, Sainsbury P, Morrissey J, Towers J, Scrivener J (1961) The reliability of psychiatric assessment: an analysis. J Ment Sci 107:887–908

Krüger G, Maurer K, Biehl H, Jung E, Schubart C (1985) 5-Jahres-Verlauf eines Samples ersterkrankter Schizophrener – Symptomatologie –. In: Reimer F (ed) Prognose und Prophylaxe in der Psychiatrie, 16. Weinsberger Kolloquium. Weissenhof, Weinsberg

Krüger G, Biehl H, Maurer K, Jung E, Bauer-Schubart C (1988) Longitudinale Symptomveränderungen bei schizophrenen Patienten über 5 Jahre nach Krankheitsbeginn. In: Bender W, Dencker SJ, Kulhanek F (eds) Schizophrene Erkrankungen. Therapie, Therapieresistenz – eine Standortbestimmung. Vieweg, Braunschweig/Wiesbaden

Olbrich R, Maurer K (1988) Zur 10. Revision der Internationalen Klassifikation der Krankheiten (ICD). Nervenheilkunde 7:139–141

Philipp M, Maier W (1987) Diagnosesysteme endogener Depressionen. Springer, Berlin Heidelberg New York

Sartorius N, Jablensky A, Cooper JE, Burke JD (1988) Psychiatric classification in an international perspective with special reference to chapter V (F) of the 10th Revision of the International Classification of Diseases "Mental, Behavioral and Development Disorders". J Psychiatry [Suppl 1] 152: 3–52

Shepherd M, Brooke EM, Cooper JE, Lin T (1968) An experimental approach to psychiatric diagnosis. An international study. Acta Psychiatr Scand [Suppl] 201:7–89

Spitzer RL, Fleiss JL (1974) A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry 125:341–347

Spitzer RL, Endicott J, Robins E (1975) Research diagnostic criteria for a selected group of functional psychoses, 2nd edn. Biometrics Research Division, New York

Stede M (1984) Einführung in die künstliche Intelligenz. Luther, Sprendlingen

Waltz DL (1982) Künstliche Intelligenz. Spektrum der Wissenschaft, December 1982, pp 68–87

Wing JK (1984) Defining the data base. World Psychiatric Association, Regional Symposium, Helsinki, June 18–21, 1984. Book of Abstracts, p 10

Wing JK, Cooper JE, Sartorius N (1974) Measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablensky A, Regier D, Sartorius N (1989) SCAN: Schedules for clinical assessment in neuropsychiatry. Arch Gen Psychiatry (in press)

World Health Organization (1988) Provisional list of categories for chapter 5 ("F") Mental Disorders in the tenth revision of the International Classification of Diseases. WHO, Geneva

Received February 13, 1989