

# Die diagnostische Übereinstimmung zwischen Klinikern und dem DIAL-Programm

## Eine Untersuchung zur computerunterstützten psychiatrischen Diagnostik

Wilhelm Schmid, Rolf Castell, Werner Mombour,  
Dieter Mittelsten Scheid und Detlev von Zerssen

Max-Planck-Institut für Psychiatrie, München

Eingegangen am 26. November 1973

### Diagnostic Agreement between Clinicians and the Program DIAL

#### A Study in Computer Assisted Diagnosis

*Summary.* DIAL, a modified version of the computer program DIAGNO II by Spitzer and Endicott, and the "Current and Past Psychopathology Scales" (CAPPS) developed for psychiatric interviews by the same authors, were tested for their diagnostic applicability in German psychiatry. The agreement of the case-history diagnoses of 98 inpatients and the diagnoses made independently by two clinicians and by the computer on the basis of the CAPPS show that the procedure can also be successfully applied in this country. The conformity of diagnoses could doubtlessly be increased still further by improvement of the scales (CAPPS) and by further modification of the diagnostic procedure.

*Key words:* Computer Assisted Diagnosis — Psychiatric Rating Scales — Diagnostic Agreement.

*Zusammenfassung.* DIAL, eine modifizierte Version des Computerprogramms DIAGNO II von Spitzer u. Endicott, und die "Current and Past Psychopathology Scales" (CAPPS), ein von denselben Autoren entwickelter Fragebogen für die psychiatrische Exploration, wurden auf ihre diagnostische Verwendbarkeit in der deutschen Psychiatrie geprüft. Der Grad der Übereinstimmung zwischen der Krankenblattdiagnose von 98 Klinikpatienten und der Diagnosen, die auf der Grundlage der CAPPS von zwei Klinikern und vom Computer jeweils unabhängig voneinander gestellt wurden, zeigt, daß das Verfahren auch hierzulande erfolgreich angewendet werden kann. Durch Verbesserung der Eingabeskalen und weitere Modifikationen des Entscheidungsprozesses könnte die Übereinstimmung zweifellos noch erhöht werden.

*Schlüsselwörter:* Computerunterstützte Diagnostik — Psychiatrische Schätzskalen — Diagnostische Übereinstimmung.

### 1. Einführung

Die Kommunikationsschwierigkeiten und die mangelhafte Übereinstimmung in der psychiatrischen Diagnostik (v. Zerssen, 1973) sind nach einer internationalen Studie von Ward *et al.* (1962) zu mehr als 60% auf Unterschiede in den nosologischen Schemata (vgl. Meyer, 1961; Mom-

bour, im Druck) zurückzuführen. Einen ersten Schritt zur Überwindung dieser „nosographischen Sprachverwirrung“ (Gross, 1969) stellt die Einigung auf den psychiatrischen Diagnoseschlüssel der WHO (ICD = International Classification of Diseases) dar, dessen 8. Revision (dtsh. Ausgabe von Degkwitz *et al.*, 1973) seit einigen Jahren in allen der WHO angeschlossenen Ländern für offizielle Gesundheitsstatistiken gültig ist. Ein weiterer Schritt in Richtung auf das Ideal einer weltweiten Übereinstimmung in der Verwendung diagnostischer Begriffe ist deren Operationalisierung durch ein allgemein verbindliches Glossar. Ein solches wurde für den psychiatrischen Teil der ICD von einem Expertenteam der WHO erarbeitet (vgl. Degkwitz *et al.*, 1973). Das Glossar der ICD enthält jedoch keine strikten Regeln zur Standardisierung des diagnostischen Vorgehens, so daß trotz seiner Verwendung die Diagnostikerübereinstimmung oft gering ist (vgl. Krauß, in Vorb.). Die bei Verwendung ähnlicher Schemata festgestellte Übereinstimmung ist nach Zubin (1967) unterschiedlich: für depressive Neurose 18–63%, für Psychopathie 6–66%. Daran sind nicht nur die in der Psychiatrie häufigen „atypischen“ und „Grenzfälle“ schuld, sondern nachgewiesenermaßen auch der Einfluß der Persönlichkeitsstruktur des Untersuchers und seiner Haltung dem Patienten gegenüber, vor allem jedoch die Anwendung unterschiedlicher Explorationstechniken und verschiedenartiger Strategien im diagnostischen Entscheidungsprozeß (Beck, 1962; Beck *et al.*, 1962; Cooper *et al.*, 1972; Gauron *et al.*, 1966; Shepherd *et al.*, 1968).

Eine Möglichkeit, den Vorgang der Diagnosenstellung zu standardisieren, ist in den Computerprogrammen DIAGNO I und DIAGNO II von Spitzer u. Endicott (1968, 1969) und in der hier untersuchten Version DIAL (Gerster u. Dirlich, in Vorb.) verwirklicht. Die Regeln für die standardisierte Informationsverarbeitung (Algorithmen; s. Lange, 1971) lassen sich beim Programmieren als einzelne logische und rechnerische Schritte (Operationen), z.B. in FORTRAN codieren und in sog. Flußdiagrammen (flow charts) schematisch darstellen.

Die Entscheidungslogik von DIAGNO basiert auf der diagnostischen Einteilung (DSM I) der American Psychiatric Association (1952): Zuerst werden Symptome abgefragt, die auf eine organisch bedingte Hirnfunktionsstörung hindeuten (Alkohol- und Drogenabusus, Cerebralsklerose). Falls dafür kein Anhalt vorliegt, wird sukzessive weitergesucht nach Hinweisen für endogene Psychosen, für vorübergehende situative Reaktionen, Störungen des Sozialverhaltens, andere Persönlichkeitsstörungen, Neurosen, psychosomatische Erkrankungen, geistigen Entwicklungsrückstand und schließlich für unspezifische Erkrankung. Das Ergebnis jeder Frage bestimmt die nächstfolgende und scheidet dabei u.U. eine oder mehrere Diagnosen oder Diagnosegruppen aus. Innerhalb jeder der großen Gruppen wird dann weiter differenziert, gegebenenfalls auch zu einer bereits ausgeschlossenen Gruppe zurückgesprungen. Das Programm hat 57 Entscheidungspunkte und ist in FORTRAN IV für die IBM 360 geschrieben.

Die Standardisierung der Informationsverarbeitung setzt aber eine standardisierte Informationsgewinnung voraus. Eine Vielzahl von genormten Untersuchungsinstrumenten wurde für diesen Zweck entwickelt (vgl. Mombour, 1972). Im einfachsten Fall hat der Untersucher im Anschluß an die Exploration eines Patienten eine Symptomenliste abzuheften. Differenzierter sind Schätzskalen, in denen die einzelnen „Items“ (in der Regel Symptome) abgestuft nach ihrem Ausprägungsgrad mit einem Punktwert (score) versehen werden. Einen hohen Grad an Objektivität gewährleisten Fragebogen mit standardisiertem Interview, in denen Reihenfolge und Formulierung der an den Patienten zu stellenden Fragen vorgeschrieben sind.

Für das Programm DIAGNO II entwickelten Spitzer u. Endicott die „Current and Past Psychopathology Scales“ (CAPPS), indem sie eine erweiterte Form der „Psychiatric Evaluation Form“ (PEF) zur Erfassung des gegenwärtigen Zustandes eines Patienten mit dem „Psychiatric History Schedule“ (PHS), das zur Dokumentation der Vorgeschichte dient, kombinierten.

Der Fragebogen besteht aus 130 Items in meist sechsfacher Abstufung über die Vorgeschichte vom 12. Lebensjahr an (Anamnese) und 41 Angaben über den Befund zum Zeitpunkt der Exploration und im vorangehenden Monat (Befund); für die vorliegende Untersuchung wurde dieser Zeitraum auf 3 Monate verlängert.

Ein Computer vermag mit dem DIAGNO II-Programm aufgrund von 94 der 171 Items der CAPPS (plus Alter und Geschlecht) eine von 44 APA-Diagnosen<sup>1</sup> zu stellen, außerdem die Diagnosen: „nicht krank“ und „unspezifische Krankheit mit leichter Symptomatik“. Die Diagnosen der APA entsprechen weitgehend den ICD-Diagnosen; der Hauptunterschied liegt darin, daß im Diagnosenschema der APA organische Psychosen in akute und chronische Formen aufgeteilt werden.

Als 1970 an der Psychiatrischen Abteilung des Max-Planck-Instituts für Psychiatrie in München Versuche zur computerunterstützten psychiatrischen Diagnostik anliefen, waren die DIAGNO-Programme die einzigen, über die in der Literatur detaillierte Angaben vorlagen. Es wurde deshalb geprüft, inwieweit das entwickeltere dieser beiden Programme (DIAGNO II) auf unsere diagnostischen Vorstellungen übertragbar wäre und ob es zu Ausbildungszwecken und in Fallkonferenzen verwendet werden könnte. Dazu war es notwendig, die diagnostische Übereinstimmung des Computers mit Klinikern, die den jeweiligen Patienten selbst exploriert hatten oder denen Symptomlisten über den betreffenden Fall vorgelegt wurden, „off-line“ zu überprüfen (vgl. Schmid, in Vorb.). Ein „on-line“-Programm (DIAL) in Form eines Frage-Antwort-Dialogs zwischen

1 APA = American Psychiatric Association.

Kliniker und Computer, in dem der Entscheidungsprozeß vom Kliniker verfolgt und korrigiert werden kann, wurde im Rahmen dieser Bemühungen um eine Standardisierung diagnostischer Prozeduren am Max-Planck-Institut für Psychiatrie, München, von Gerster u. Dirlich (im Druck) aus dem DIAGNO II-Programm entwickelt. Das DIAL-Programm ist so allgemein gefaßt, daß es auf beliebige Skalen und beliebige Entscheidungsbäume anwendbar ist. Seine erste Anwendung bezog sich aber auf die CAPPS und den nur geringfügig modifizierten<sup>2</sup> Entscheidungsprozeß des DIAGNO II-Programms. Dieses modifizierte Programm wurde in der vorliegenden Untersuchung im "off-line"-Verfahren verwendet.

## 2. Methodik

Wir verglichen die Diagnosen, die an denselben Fällen von verschiedenen „Diagnostikern“ gestellt wurden. Bei den Fällen handelte es sich um 98 Patienten (2 von ursprünglich 100 Krankengeschichten wurden aus technischen Gründen ausgeschieden), die 1969/1970 auf der psychiatrischen Abteilung des MPI stationär behandelt worden waren. Die Fälle waren nach folgenden Gesichtspunkten ausgewählt worden: Erstens sollte die vorläufige Diagnose, die im Anschluß an das Aufnahmeinterview gestellt worden war, mit der Entlassungsdiagnose übereinstimmen. Unklare Fälle oder Patienten, bei denen während des stationären Aufenthalts die Diagnose revidiert werden mußte, wurden ausgeschieden. Zweitens sollten möglichst alle der im DIAL-Programm vorgesehenen Diagnosen vertreten sein. Die in den Krankengeschichten und zusätzlichen Befundbogen (s.u.) enthaltene Information wurde von einem der Autoren (S.) in die CAPPS übertragen.

Die Daten für den Anamnese-Teil (A) der CAPPS wurden den Krankengeschichten entnommen, die von 21 verschiedenen Klinikern verfaßt und folgendermaßen aufgebaut waren: Familienanamnese, äußere Lebensgeschichte, frühere Erkrankungen, innere Lebensgeschichte, Vorgeschichte der jetzigen Erkrankung, psychischer Befund, Diagnose, Differentialdiagnose, weitere diagnostische Maßnahmen und Behandlungsplan. Die psychopathologischen Daten für den Befund-Teil (B) der CAPPS entstammen im wesentlichen zwei standardisierten Fragebogen zur Dokumentation des psychopathologischen Befundes und zusätzlichen Angaben aus den

2 Dabei sind die Diagnosen des Programms dem ICD-Schema angepaßt und ihre Anzahl ist auf 50 erhöht. Außerdem wird im Knoten 120 für die Entscheidung, ob weiter in Richtung auf endogene Psychose oder aber auf Neurose hin abgefragt wird, der Schwellenwert angehoben. In diesem Knoten wird nach dem Vorhandensein von Wahnideen und Halluzinationen gefragt. War nur eines dieser beiden Symptome „angedeutet vorhanden“ (Ausprägungsgrad 2), so wurde — vor unserer Änderung — bereits eine endogene Psychose diagnostiziert. Da auch bei neurotischen Erkrankungen paranoide Symptome „angedeutet vorhanden“ sein können, haben wir den Schwellenwert für die Summe der Werte beider Symptome von 2 auf 4 erhöht.

Krankengeschichten. Die Fragebogen waren die IMPS (Inpatient Multidimensional Psychiatric Scale) von Lorr *et al.* (1962) und Blatt 3 („Psychischer Befund“) des AMP-Systems (Scharfetter, 1972). Gelegentlich bestanden erhebliche Abweichungen zwischen den Angaben der IMPS und der AMP-Skalen zum selben Patienten. In diesen Fällen wurde in der Regel diejenige Angabe berücksichtigt, die das betreffende Item als vorhanden bzw. in stärkerer Ausprägung dokumentierte.

Die „Diagnostiker“, deren Beurteilungen miteinander verglichen wurden, waren

a) diejenigen Kliniker, die den betreffenden Patienten während seines Klinikaufenthaltes untersucht und die Diagnose in der Krankengeschichte (KG) niedergelegt haben;

b) zwei „Fragebogendiagnostiker“, d.h. zwei Kliniker, die nur mit der Information aus den CAPPs, also ohne persönliche Kenntnis der Patienten, auskommen mußten (FD1 und FD2).

FD1 ist Facharzt für Psychiatrie mit langjähriger klinischer Erfahrung und seit 1966 an der Abteilung für Erwachsenenpsychiatrie am Max-Planck-Institut für Psychiatrie tätig. FD2 — ebenfalls Facharzt für Psychiatrie — war z.Z. der Untersuchung 4 Jahre am Max-Planck-Institut für Psychiatrie, davon 1 Jahr an der Abteilung für Erwachsenenpsychiatrie, tätig gewesen. FD1 und FD2 stellten anhand der CAPPs-Listen unabhängig voneinander ihre Diagnosen.

c) ein Computer, dem die gleiche standardisierte Information über den Patienten zur Verfügung stand und der die diagnostische Entscheidung mit Hilfe des DIAL-Programms durchführte. Er wählte aus 50 in dem Programm vorgegebenen ICD-Diagnosen jeweils eine Haupt- und ggf. eine Zusatzdiagnose aus.

Zur statistischen Auswertung wurden die 50 möglichen DIAL-Diagnosen zu Gruppen zusammengefaßt, die in drei aufeinanderfolgenden Ansätzen immer weiter aufgefächert wurden. In der ersten Einteilung wurde zunächst grob unterschieden zwischen Psychose und Nicht-Psychose (Neurose, Psychopathie usw.). Psychotische und neurotische Depression wurden dabei ausgeklammert und mit „reaktiver depressiver Psychose“ (ICD-Nr. 298.0) zusammengefaßt zu einer dritten Gruppe „Depression“. Die restlichen Kategorien waren Oligophrenie, „nicht psychisch krank“ und „unklarer Fall“.

In der zweiten Einteilung wurden die Psychosen in körperlich begründbare und endogene aufgeteilt und die Süchte von den Neurosen und den übrigen nichtpsychotischen Störungen abgetrennt. Die anderen Gruppen blieben gleich.

Die dritte Einteilung differenzierte genauer und unterschied die folgenden Kategorien: Alkoholpsychosen/sonstige Psychosen bei organischen Erkrankungen/Schizophrenie/schizoaffektive Psychose/Manie/Depression/Neurosen ohne Depression/Persönlichkeitsstörungen/sexuelle Verhaltensabweichungen/Alkoholismus/sonstige Süchte/psychosomati-

sche Erkrankungen/nicht-psychotische psychische Störungen bei körperlichen Erkrankungen/Oligophrenie/nicht psychisch krank/unklarer Fall.

Von den vier Diagnosen, die zu jedem Patienten vorlagen, wurde zunächst jede mit jeder anderen verglichen und ihre Übereinstimmung innerhalb der drei Einteilungen festgestellt. Sodann untersuchten wir, wie sich der Übereinstimmungsgrad bei den einzelnen diagnostischen Kategorien unterschied. Danach wurden neben den Erstdiagnosen auch noch die Zweitdiagnosen berücksichtigt.

Die Übereinstimmung wurde jeweils in Prozentwerten ausgedrückt. Die Verwendung eines speziellen Übereinstimmungskoeffizienten, z.B.  $\kappa$  (Kappa; vgl. Spitzer *et al.*, 1967) erschien uns für unsere Zwecke keinen Vorteil zu bringen, da es uns in dieser Untersuchung nur auf einen zahlenmäßigen Vergleich der paarweisen Übereinstimmung ankam.

### 3. Ergebnisse und Diskussion

Die Übereinstimmung der Erstdiagnosen von Krankengeschichte, DIAL, FD1 und FD2 ist in Tab. 1 dargestellt. Man sieht, daß in allen drei Einteilungen DIAL mit der Diagnose laut Krankengeschichte nahezu so gut wie FD2 und sogar besser als FD1 übereinstimmt, d.h. der Computer leistet dasselbe wie ein Fragebogendiagnostiker, dem die gleiche Information zur Verfügung steht. Daß die beiden Fragebogendiagnostiker, die am selben Institut tätig sind, untereinander am besten übereinstimmen würden, war zu erwarten. Ihnen stand die gleiche Information zur Verfügung, wogegen sich die Krankengeschichtsdiagnose auch auf die Information aus dem persönlichen Kontakt mehrerer Ärzte mit dem Patienten und seinen Angehörigen, also andersartiges und umfangreicheres Material stützte. Der Computer dagegen arbeitete mit derselben beschränkten Information wie die FD, aber mit einem Programm, das von Fachleuten mit einer anderen Ausbildung auf einem anderen Kontinent entwickelt wurde.

Der Anteil übereinstimmend gewählter Kategorien wird mit zunehmender Differenzierung immer kleiner und ist bei Unterscheidung von effektiv 14 psychopathologischen Kategorien (Einteilung 3) um 15–20% niedriger als bei Unterscheidung von nur vier Krankheitskategorien (Einteilung 1). Dies ist ebenfalls zu erwarten und wird grundsätzlich von allen Reliabilitätsprüfungen psychiatrischer Diagnosen berichtet (Zigler u. Philips, 1961). Die Übereinstimmung mit den Computerdiagnosen liegt dabei in Einteilung 3 um 2–7% unter dem Gesamtdurchschnitt.

In der Annahme, daß bestimmte Diagnosen leicht und einstimmig gestellt werden, andere dagegen mehr Schwierigkeiten bereiten und abweichend gewählt werden, untersuchten wir als nächstes die Übereinstimmung innerhalb der diagnostischen Kategorien. Dabei wurden wieder nur die Hauptdiagnosen berücksichtigt (Tab. 2–4).

Tabelle 1. Übereinstimmung der Erstdiagnosen. KG = Diagnose laut Krankengeschichte; DIAL = Computordiagnose nach dem DIAL-Programm; FD1 und FD2 = Diagnosen der beiden „Fragebogendiagnostiker“

|              | KG-FD1<br>o/o | KG-FD2<br>o/o | KG-DIAL<br>o/o | FD1-DIAL<br>o/o | FD2-DIAL<br>o/o | FD1-FD2<br>o/o | Total<br>o/o |
|--------------|---------------|---------------|----------------|-----------------|-----------------|----------------|--------------|
| Einteilung 1 | 73            | 78            | 74             | 67              | 70              | 76             | 73           |
| Einteilung 2 | 61            | 69            | 65             | 61              | 66              | 70             | 65           |
| Einteilung 3 | 54            | 62            | 54             | 49              | 53              | 64             | 56           |

Tabelle 2. Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 1. Zu jedem Diagnostikerpaar ist angegeben, wie oft jeder von ihnen die betreffende diagnostische Gruppe wählte; dazwischen in Klammern die Anzahl der Fälle, in denen diese Kategorie übereinstimmend von beiden gewählt wurde. Darunter Übereinstimmung in Prozent, bezogen auf die Häufigkeit, mit der die Kategorie von mindestens einem der beiden Diagnostiker gewählt wurde

|               | KG-FD1            | KG-FD2            | KG-DIAL           | FD1-DIAL          | FD2-DIAL          | FD1-FD2           | Total            |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|
| Psychosen     | 41 (29) 37<br>59% | 41 (35) 44<br>70% | 41 (35) 46<br>67% | 37 (31) 46<br>60% | 44 (36) 46<br>67% | 37 (31) 44<br>62% | 307 (197)<br>64% |
| Depression    | 18 (13) 25<br>43% | 18 (14) 21<br>56% | 18 (9) 11<br>45%  | 25 (8) 11<br>29%  | 21 (9) 11<br>39%  | 25 (16) 21<br>53% | 156 (69)<br>44%  |
| Neurosen usw. | 39 (30) 36<br>67% | 39 (27) 33<br>60% | 39 (29) 37<br>62% | 36 (27) 37<br>59% | 33 (24) 37<br>52% | 36 (27) 33<br>64% | 271 (164)<br>61% |
| Unklar        | —                 | —                 | 0 (0) 4<br>0%     | 0 (0) 4<br>0%     | 0 (0) 4<br>0%     | —                 | 12 (0)<br>0%     |

Tabelle 3. Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 2. Erläuterungen vgl. Tab. 2

|                      | KG-FD1            | KG-FD2            | KG-DIAL           | FD1-DIAL          | FD2-DIAL          | FD1-FD2           | Total            |
|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|
| Organische Psychosen | 11 (5) 14<br>25%  | 11 (5) 11<br>29%  | 11 (3) 8<br>19%   | 14 (6) 8<br>38%   | 11 (5) 8<br>36%   | 14 (9) 11<br>56%  | 99 (33)<br>33%   |
| Endogene Psychosen   | 30 (18) 23<br>51% | 30 (25) 33<br>66% | 30 (25) 38<br>58% | 23 (21) 38<br>53% | 33 (27) 38<br>61% | 23 (19) 33<br>51% | 237 (135)<br>57% |
| Depression           | 18 (13) 25<br>43% | 18 (14) 21<br>56% | 18 (9) 11<br>45%  | 25 (8) 11<br>29%  | 21 (9) 11<br>39%  | 25 (16) 21<br>53% | 156 (69)<br>44%  |
| Neurosen usw.        | 32 (17) 20<br>49% | 32 (18) 22<br>50% | 32 (20) 27<br>51% | 20 (15) 27<br>47% | 22 (16) 27<br>48% | 20 (15) 22<br>56% | 202 (101)<br>50% |
| Süchte               | 7 (7) 16<br>44%   | 7 (6) 11<br>50%   | 7 (7) 10<br>70%   | 16 (10) 10<br>63% | 11 (8) 10<br>62%  | 16 (10) 11<br>59% | 84 (48)<br>57%   |
| Unklar               | —                 | —                 | 0 (0) 4<br>0%     | 0 (0) 4<br>0%     | 0 (0) 4<br>0%     | —                 | 12 (0)<br>0%     |



Tabelle 4. Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 3. Erläuterungen vgl. Tab. 2

|   | KG-FD 1        | KG-FD 2        | KG-DIAL        | FD1-DIAL       | FD2-DIAL       | FD1-FD2        | Total            |
|---|----------------|----------------|----------------|----------------|----------------|----------------|------------------|
| Alkoholpsychosen                        | 1 (1)<br>17%   | 1 (1)<br>20%   | 1 (0)<br>0%    | 4 (3)<br>43%   | 5 (3)<br>50%   | 6 (4)<br>57%   | 36 (12)<br>33%   |
| Organische Psychosen                    | 10 (3)<br>20%  | 10 (3)<br>23%  | 10 (3)<br>27%  | 4 (2)<br>20%   | 4 (2)<br>25%   | 8 (3)<br>27%   | 68 (16)<br>24%   |
| Schizophrenie                           | 26 (15)<br>50% | 26 (20)<br>59% | 26 (23)<br>61% | 35 (18)<br>50% | 35 (20)<br>47% | 28 (14)<br>42% | 214 (110)<br>51% |
| Schizoaffektive Psychose                | 3 (0)<br>0%    | 3 (0)<br>0%    | 3 (0)<br>0%    | 2 (0)<br>0%    | 2 (0)<br>0%    | 1 (0)<br>0%    | 3 (0)<br>0%      |
| Manie                                   | 1 (1)<br>50%   | 1 (1)<br>50%   | 1 (0)<br>0%    | 1 (1)<br>50%   | 1 (1)<br>50%   | 2 (2)<br>100%  | 12 (6)<br>50%    |
| Depression                              | 18 (13)<br>43% | 18 (14)<br>56% | 18 (9)<br>45%  | 11 (8)<br>29%  | 11 (9)<br>39%  | 25 (16)<br>53% | 156 (69)<br>44%  |
| Neurosen                                | 16 (12)<br>63% | 16 (13)<br>76% | 16 (7)<br>33%  | 12 (4)<br>17%  | 12 (5)<br>24%  | 15 (11)<br>61% | 119 (52)<br>44%  |
| Psychopathien                           | 6 (0)<br>0%    | 6 (1)<br>17%   | 6 (1)<br>11%   | 4 (0)<br>0%    | 4 (1)<br>25%   | 1 (0)<br>0%    | 33 (3)<br>9%     |
| Sexuelle Abweichungen                   | 2 (2)<br>100%  | 2 (2)<br>67%   | 2 (2)<br>100%  | 2 (2)<br>100%  | 2 (2)<br>67%   | 3 (2)<br>67%   | 15 (12)<br>80%   |
| Alkoholabusus                           | 4 (3)<br>33%   | 4 (2)<br>33%   | 4 (3)<br>60%   | 4 (4)<br>50%   | 4 (3)<br>60%   | 8 (3)<br>33%   | 42 (18)<br>43%   |
| Sonstige Süchte                         | 3 (3)<br>38%   | 3 (3)<br>43%   | 3 (3)<br>50%   | 6 (6)<br>75%   | 6 (5)<br>63%   | 8 (7)<br>88%   | 45 (27)<br>60%   |
| Psychosomatische Krankheiten            | 3 (0)<br>0%    | 3 (0)<br>0%    | 3 (1)<br>13%   | 6 (0)<br>0%    | 6 (0)<br>0%    | 0 (0)<br>—     | 26 (1)<br>4%     |
| Psychische Störung bei org. Krankheiten | 5 (0)<br>0%    | 5 (1)<br>13%   | 5 (1)<br>14%   | 3 (0)<br>0%    | 3 (1)<br>17%   | 2 (0)<br>0%    | 39 (3)<br>8%     |
| Unklar                                  | —              | —              | 0 (0)<br>0%    | 4 (0)<br>0%    | 4 (0)<br>0%    | —              | 12 (0)<br>0%     |

Relativ am besten ist die Gesamtübereinstimmung bei Schizophrenen (durchschnittlich 51%); für Süchte (60%) variiert die paarweise Übereinstimmung beträchtlich (zwischen 38 und 88%), dasselbe gilt für Neurosen (17–76%; durchschnittlich 44%). Die hohe Übereinstimmung bei sexuellen Deviationen (80%) und bei Manie (50%) sagt wegen des engen Zusammenhangs zwischen Leitsymptom und Diagnose und wegen der geringeren Anzahl der Fälle nicht viel aus. Letzteres gilt auch für die Nichtübereinstimmung bei den schizoaffektiven Psychosen (0%).

In der Kategorie „Depression“ wurden die Diagnosen psychotische und neurotische Depression mit reaktiver depressiver Psychose zusammengefaßt (Übereinstimmung 29–56%; durchschnittlich 44%); die Abgrenzung dieser Kategorie gelang nicht befriedigend. Viermal diagnostizierte der Computer eine Angstneurose in Fällen, wo Krankengeschichte und Fragebogendiagnostiker übereinstimmend eine Depression diagnostizierten. (Bei der Faktorenanalyse der CAPPS bilden Angst und Depression einen gemeinsamen Faktor: Spitzer u. Endicott, 1969.) Eine Untersuchung von Sandifer (1964), in der zehn Teilnehmer einer Fallkonferenz jeweils anschließend den betreffenden Patienten einer von 12 diagnostischen Kategorien zuzuordnen hatten, ergab bei affektiven Störungen ebenfalls sehr niedrige Übereinstimmung (Involutionspsychose 26%, manisch-depressive Psychose 36%).

Bei psychosomatischen Störungen und Psychopathien herrscht praktisch keine Übereinstimmung (4 bzw. 9%). Die für die Diagnostik von Psychopathien notwendigen fremdanamnestischen Angaben, die der Kliniker aus dem Gespräch mit Angehörigen und Bekannten des Patienten erhält, werden im CAPPS-Bogen nicht ausreichend festgehalten. Für die beiden Fragebogendiagnostiker bedeutete die Beschränkung auf den CAPPS-Bogen außerdem den Wegfall der „nichtverbalisierten“ Information, die der Kliniker beim persönlichen Interview gewinnt. Wie ungern sich Psychiater auf die schriftlich festgelegten Beobachtungen und Interpretationen ihrer Kollegen verlassen, beschreibt Gauron (1966) in seiner Untersuchung des diagnostischen Vorgehens.

Die Nichtübereinstimmung für nicht-psychotische psychische Störungen bei körperlichen Erkrankungen korrespondiert mit der geringen Übereinstimmung bei organischen Psychosen (durchschnittlich 24%): Häufig wird derselbe Patient z.B. von FD2 als psychisch gestörter, aber nicht-psychotischer „Organiker“ und von FD1 als psychotischer „Organiker“ eingestuft.

Schließlich wurden auch die Zweitdiagnosen verwendet, und zwar auf folgende Weise: Wenn beide Diagnostiker, etwa FD1 und FD2, nur je eine Diagnose gestellt hatten, kann nur ein Diagnosenpaar verglichen werden. Wenn beide eine Erst- und eine Zweitdiagnose angegeben haben, haben wir zwei Paare gebildet. In Fällen, wo

Tabelle 5. Übereinstimmung von Haupt- und Zusatzdiagnosen. Angegeben ist jeweils die Anzahl möglicher Diagnosenpaare, dahinter in Klammern die Anzahl übereinstimmender Diagnosenpaare. Darunter Übereinstimmung in Prozent

|              | KG-FD1          | KG-FD2           | KG-DIAL         | FD1-DIAL        | FD2-DIAL         | FD1-FD2         |
|--------------|-----------------|------------------|-----------------|-----------------|------------------|-----------------|
| Einteilung 1 | 140 (81)<br>58% | 169 (106)<br>63% | 171 (96)<br>56% | 136 (72)<br>53% | 179 (100)<br>56% | 145 (83)<br>57% |
| Einteilung 2 | 140 (69)<br>49% | 169 (100)<br>59% | 171 (88)<br>51% | 136 (65)<br>48% | 179 (89)<br>50%  | 145 (77)<br>53% |
| Einteilung 3 | 140 (61)<br>44% | 169 (82)<br>48%  | 171 (68)<br>40% | 136 (50)<br>37% | 179 (66)<br>37%  | 145 (69)<br>48% |

FD1 jedoch zwei und FD2 nur eine Diagnose gestellt hat, gingen wir ebenfalls von zwei Paaren aus; denn die einzige Diagnose von FD2 kann mit jeder der beiden Diagnosen von FD1 verglichen werden. Außerdem ist eine nicht gestellte Zweitdiagnose in Fällen, wo der andere Diagnostiker eine Zusatzdiagnose für zutreffend erachtet hat, auch als eine Art „nicht übereinstimmender“ Diagnose anzusehen. Bei der Auszählung der übereinstimmenden Diagnosenpaare wird jede einzelne Diagnose öfter als einmal berücksichtigt. Durch diese Methode fielen die Werte niedriger aus als beim Vergleich der Erstdiagnosen (s. Tab. 5).

Ein weiterer Weg, die Übereinstimmung aller Diagnostiker auszudrücken, wäre folgender: Zunächst wird jede der möglichen Diagnosen mit jeder anderen verglichen und ein Zahlenwert für den Grad ihrer „Verschiedenheit“ (disagreement level) festgesetzt. Dies ist zwangsläufig ein subjektives Vorgehen und sollte von mehreren erfahrenen Klinikern durchgeführt werden. Beim Vergleich zwischen zwei Diagnostikern wird bei „ähnlichen“ Diagnosen die Teilübereinstimmung berücksichtigt, was im allgemeinen eine höherprozentige Übereinstimmung ergibt als das von uns verwendete Verfahren (vgl. Schmid, in Vorb.); so stimmt bei Spitzer (1969) zum erstenmal in der Literatur der Computer mit Klinikern ebenso gut überein wie diese untereinander, wenn allen dieselbe schriftliche Information zur Verfügung steht. Psychiater, die den betreffenden Patienten persönlich genau kennen, stimmen mit dem Computer weniger gut überein.

Die Abhängigkeit der diagnostischen Übereinstimmung vom Sicherheitsgrad, mit dem die jeweilige Diagnose gestellt wurde, wurde von uns nicht berücksichtigt. Beck *et al.* (1962) berichten, daß die Übereinstimmung zwischen Klinikern, die unabhängig voneinander dieselben Fälle explorierten und diagnostizierten, von 81% bei „Sicherheit“ auf unter 50% bei „Wahrscheinlichkeit“ und „Unsicherheit“ der Diagnose absank. Sandifer (1964) beschreibt den Effekt als weniger gravierend (Absinken von 65 auf 45%).

#### 4. Folgerung

Beim Vergleich zwischen der Computerdiagnose und der klinischen Diagnose aus der Krankengeschichte, die durch ein Ärzteteam gestellt wurde, erwies sich das modifizierte DIAGNO II-Programm als brauchbar. Dabei arbeitete der Computer in dieser Untersuchung mit Informationen, die der Krankengeschichte des Patienten, AMP-Skalen und Lorr-Skalen entnommen waren. Es ist anzunehmen, daß sich die diagnostische Übereinstimmung erhöhen würde, wenn die CAPPS vom jeweilig explorierenden Arzt selbst ausgefüllt worden wäre, da seine diagnostische Präferenz das Symptomprofil „überzeichnen“ dürfte. Auch ist der Fragebogen in Einzelheiten noch verbesserungsfähig (zur Kritik an den CAPPS vgl. Schmid, in Vorb.). So wäre ein für Ausbildungszwecke ausreichender Grad an diagnostischer Zuverlässigkeit zu erreichen. Jedoch sollte man sich nicht starr an das von uns verwendete Computerprogramm halten, sondern vorläufig noch weiter damit experimentieren. Das hier dargestellte Prinzip der Computerdiagnostik ist zudem ein Verfahren, um Verschiedenheiten von Diagnostikern zu erhellen. Für die psychiatrische Ausbildung erscheint der bei DIAL mögliche Dialog zwischen Arzt und Computer (manmachine-dialogue) besonders geeignet (v. Zerssen, 1972).

#### Literatur

- American Psychiatric Association: Diagnostic and statistical manual: mental disorders. Washington: American Psychiatric Association 1952
- Beck, A. T.: Reliability of psychiatric diagnosis: 1. A critique of systematic studies. *Amer. J. Psychiat.* **119**, 210—216 (1962)
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., Erbaugh, J. K.: Reliability of psychiatric diagnosis: A study of consistency of clinical judgments and ratings. *Amer. J. Psychiat.* **119**, 351—357 (1962)
- Cooper, J. E., Kendell, R. E., Gurland, B. J., Sharpe, L., Copeland, J. R. M., Simon, R.: Psychiatric diagnosis in New York and London. London-New York-Toronto: Oxford University Press 1972
- Degkwitz, R., Helmchen, H., Mombour, W. (Hrsg.): Diagnoseschlüssel und Glossar psychiatrischer Krankheiten. Deutsche Übersetzung der internationalen Klassifikation der WHO (ICD), 8. Revision, und des internationalen Glossars. 3. Aufl. Berlin-Heidelberg-New York: Springer 1973
- Gauron, E. F., Dickinson, J. K.: Diagnostic decision making in psychiatry: I. Information usage. *Arch. gen. Psychiat.* **14**, 225—232 (1966); II. Diagnostic styles. *Arch. gen. Psychiat.* **14**, 233—237 (1966)
- Gerster, F., Dirllich, G.: DIAL, ein Programm zur Entscheidungshilfe in der Psychiatrie für das IBM-System 1130 (in Vorbereitung)
- Gross, R.: Medizinische Diagnostik — Grundlagen und Praxis. Berlin-Heidelberg-New York: Springer 1969
- Krauß, B.: Empirische Untersuchung zur Verlässlichkeit psychiatrischer Diagnosen. Med. Diss., Univ. München (in Vorbereitung)
- Lange, H.-J.: Algorithmische Diagnostik. *EDV, BMT, NM* **2**, 577—580 (1971)
- Lorr, M., Klett, C. J., McNair, D. M., Lasky, J. J.: Inpatient multidimensional psychiatric scale (IMPS). Palo Alto/Calif.: Consulting Psychologists Press 1962

- Meyer, J. E.: Diagnostische Einteilungen und Diagnosenschemata in der Psychiatrie. In: H. W. Gruhle, R. Jung, W. Mayer-Gross u. M. Müller: Psychiatrie der Gegenwart, Bd. 3, S. 130—180. Berlin-Göttingen-Heidelberg: Springer 1961
- Mombour, W.: Klassifikation, Patientenstatistik, Register. In: K. P. Kisker, J. E. Meyer, M. Müller u. E. Strömberg: Psychiatrie der Gegenwart. 2. Aufl., Bd. 3. Berlin-Heidelberg-New York: Springer (im Druck)
- Mombour, W.: Verfahren zur Standardisierung des psychopathologischen Befundes. Psychiat. Clin. **5**, 73—120, 137—157 (1972)
- Sandifer, M. G., Pettus, C., Quade, D.: A study of psychiatric diagnosis. J. nerv. ment. Dis. **139**, 350—356 (1964)
- Scharfetter, C.: Das AMP-System, 2. Aufl. Berlin-Heidelberg-New York: Springer 1972
- Schmid, W.: Untersuchung über die diagnostische Übereinstimmung zwischen Klinikern und dem System DIAL zur computerunterstützten psychiatrischen Diagnostik. Med. Diss., Univ. München (1974)
- Shepherd, M., Brooke, E. M., Cooper, J. E., Lin, T.: An experimental approach to psychiatric diagnosis. Acta psychiat. scand., Suppl. 201 ad vol. **44** (1968)
- Spitzer, R. L., Cohen, J., Fleiss, J. L., Endicott, J.: Quantification of agreement in psychiatric diagnosis: A new approach. Arch. gen. Psychiat. **17**, 83—87 (1967)
- Spitzer, R. L., Endicott, J.: DIAGNO: A computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. Arch. gen. Psychiat. **18**, 746—756 (1968)
- Spitzer, R. L., Endicott, J.: DIAGNO II: Further developments in a computer program for psychiatric diagnosis. Amer. J. Psychiat. **125**, Suppl., 12—21 (1969)
- Ward, C. H., Beck, A. T., Mendelson, M., Mock, J. E., Erbaugh, J. K.: The psychiatric nomenclature: Reasons for diagnostic disagreement. Arch. gen. Psychiat. **7**, 198—205 (1962)
- Zerssen, D. v.: Vorschläge für die klinische Ausbildung in psychiatrischer Diagnostik. 87. Wanderversammlung Südwestdeutscher Psychiater, Baden-Baden 1971. Ref.: Zbl. ges. Neurol. Psychiat. **204**, 546 (1972)
- Zerssen, D. v.: Diagnose. Nosologie. In: C. Müller: Lexikon der Psychiatrie. Berlin-Heidelberg-New York: Springer 1973
- Zigler, E., Philipps, L.: Psychiatric diagnosis: A critique. J. abnorm. soc. Psychol. **63**, 607—618 (1961)
- Zubin, J.: Classification of the behavior disorders. Ann. Rev. Psychol. **18**, 373—406 (1967)

Prof. Dr. D. v. Zerssen  
Max-Planck-Institut für Psychiatrie  
D-8000 München 40  
Kraepelinstraße 10  
Bundesrepublik Deutschland