

## Psychiatrische Klassifikation durch diskriminanzanalytische Anwendung der Q-Faktorenanalyse

W. Schubö, U. Hentschel, D. v. Zerssen und W. Mombour

Max-Planck-Institut für Psychiatrie, München

Eingegangen am 2. Mai 1975

### Psychiatric Classification by Means of a Discriminatory Application of Q Factor Analysis

*Summary.* An attempt is made to separate different pairs of psychiatric patient groups by means of a modified form of *Q* factor analysis comparable to discriminant analysis. The psychopathological state of 454 patients had been rated using two psychiatric rating scales, the IMPS (Inpatient Multidimensional Psychiatric Scale) by Lorr *et al.* and the psychopathological and somatic scales of the AMP documentation system. Out of these patients the four most frequently occurring groups (schizophrenia, paranoid form,  $n = 45$ ; schizophrenia, unspecified form,  $n = 47$ ; depressive psychosis,  $n = 44$ ; depressive neurosis,  $n = 53$ ) were selected. Each patient group was divided randomly into two samples, an analysis sample and a validation sample. Only those items were selected which discriminated best between any two analysis samples. Using *G* indices a *Q* factor analysis was calculated, and the results improved by a criterion-related additional rotation. The resulting weights were transferred to the validation samples in order to have a cross validation. The mean percentage of correct placements within the validation samples was 83%.

*Key words:* Classification by Means of Discriminant Analysis — *G* Index — *Q* Factor Analysis — Psychiatric Rating Scales — Algorithmic Diagnosis.

*Zusammenfassung.* Es wird der Versuch gemacht, mit Hilfe einer modifizierten Form der *Q*-Faktorenanalyse verschiedene psychiatrische Patientengruppen paarweise diskriminanzanalytisch voneinander zu trennen. Dem Gruppenvergleich wurden Daten des psychopathologischen Befundes zugrunde gelegt, der mit zwei psychopathologischen Schätzskalen, der Inpatient Multidimensional Psychiatric Scale (IMPS) von Lorr *et al.* und den Bögen für den psychischen und somatischen Befund des AMP-Dokumentationssystems, erfaßt worden war. Von insgesamt 454 Patienten wurden die vier größten diagnostischen Gruppen (Schizophrenie, paranoide Form,  $n = 45$ ; Schizophrenie, nicht näher bezeichnete Form,  $n = 47$ ; Depression im Rahmen einer affektiven Psychose,  $n = 44$ ; depressive Neurose,  $n = 53$ ) ausgewählt. Die Diagnosengruppen wurden jeweils in Analyse- und Validierungsstichprobe unterteilt. Mit Hilfe der Analysestichproben wurden die für die Zuordnung trennschärfsten Items ausgewählt, eine *Q*-Faktorenanalyse mit *G*-Indices gerechnet und durch eine kriteriumsbezogene Nachrotation verbessert. Die dabei erhaltenen Gewichte wurden für die Kreuzvalidierung an den Validierungsstichproben übernommen, wobei durchschnittlich 83% der Patienten richtig klassifiziert wurden.

*Schlüsselwörter:* Diskriminanzanalytische Klassifikation — *G*-Index — *Q*-Faktorenanalyse — Psychiatrische Schätzskalen — Algorithmische Diagnostik.

### A. Einleitung

In dem Bemühen, für wissenschaftliche Aufgabenstellungen das sonst weitgehend intuitive Vorgehen des Klinikers bei der psychiatrischen Fallbeurteilung zu standardisieren, wurden in den letzten Dezennien klinische Beurteilungsskalen zur Erhebung und Dokumentation des psychopathologischen Befundes entwickelt

(vgl. Mombour, 1972). Durch die Anwendung multivariater statistischer Analyseverfahren auf das mittels solcher Schätzskalen gewonnene Datenmaterial lassen sich klinische Syndrome (als Merkmals-Cluster bzw. -Faktoren) in einer kontrollierbaren Weise objektivieren (Lorr *et al.*, 1963; Mombour *et al.*, 1973). Auf dieser Basis können für den Einzelfall „Syndrom-Scores“ errechnet und Durchschnittswerte für bestimmte Gruppen von Patienten (z. B. jeweils Fälle einer diagnostischen Kategorie) ermittelt werden. Die so gewonnenen „Syndrom-Profile“ dienen der quantifizierenden Beschreibung der Psychopathologie des Einzelfalles bzw. einer bestimmten Patientengruppe (vgl. Mombour, 1974).

Die Zuordnung eines Falles zu einer Kategorie innerhalb eines nosologischen Systems — also die nosologische Diagnostik (vgl. v. Zerssen, 1973) — bleibt aber auch bei der Verwendung klinischer Schätzskalen gewöhnlich noch weitgehend der Intuition des Klinikers überlassen. Dadurch werden Vergleiche der Untersuchungsbefunde verschiedener Kliniker mit einem schwer abschätzbarer subjektiven Unsicherheitsfaktor belastet.

Um den Weg vom Befund zur Diagnose im Sinne einer „algorithmischen Diagnostik“ (Ledley, 1965; Lange, 1971) zu standardisieren, hat man Programme entwickelt (Gorry, 1967; Bégon *et al.*, 1972; Blomer *et al.*, 1972; Horrocks *et al.*, 1972), die das Vorgehen des Klinikers bei der Diagnosenstellung simulieren. Die Anwendung solcher Verfahren in der Psychiatrie setzt allerdings — zusätzlich zur Anwendung von Skalen zur Abschätzung des psychopathologischen Befundes — die standardisierte Gewinnung somatischer Befunddaten sowie anamnestischer Angaben voraus.

Man kann zwei verschiedene Prinzipien der Zuordnung von Daten aus Anamnese und Befund zu diagnostischen Kategorien unterscheiden, ein sog. deterministisches (vgl. Jesdinsky, 1973) und ein sog. probabilistisches (vgl. Victor, 1973). Bei Anwendung des deterministischen Prinzips erfolgt eine eindeutige Zuordnung aufgrund logischer Verknüpfungen, entweder in einem sukzessiven Vorgehen anhand eines Entscheidungsbaumes wie in den DIAGNO I- und II-Programmen von Spitzer und Endicott (1968, 1969) bzw. deren Modifikation im DIAL-Programm (Schmid *et al.*, 1974; Gerster u. Dirlich, 1975) oder im simultanen „Befundlistenvergleich“. Mustererkennungs-Programme, zu denen man z. B. das Catego-Programm von Wing *et al.* (1974) rechnen könnte, bilden einen Übergang zu probabilistischen Verfahrensweisen, soweit bei der Zuordnung eines Falles aufgrund der bei ihm festgestellten Merkmalskonfiguration zu den eine diagnostische Kategorie repräsentierenden „typischen“ Konfigurationen verschiedene Grade der Übereinstimmung berücksichtigt und als „Wahrscheinlichkeiten“ der Zugehörigkeit des Falles zu der betreffenden Kategorie interpretiert werden. Bei den im engeren Sinne probabilistischen Verfahren werden die Merkmale entsprechend ihrer differentialdiagnostischen Bedeutung gewichtet, bei der diagnostischen Anwendung des „Bayesschen Theorems“ auch die Häufigkeiten der in Frage kommenden Krankheitsformen (in der jeweiligen Referenzpopulation) in Rechnung gestellt. Diese Häufigkeiten können ebenso wie die Merkmalsgewichte von Experten geschätzt oder aber empirisch ermittelt werden. In der Regel werden auch die probabilistischen Modelle, die zunächst nur die Wahrscheinlichkeit angeben, mit der ein Fall verschiedenen Kategorien zugehört, deterministisch angewendet in dem Sinne, daß man zu einer eindeutigen diagnostischen Entscheidung kommt.

Ein gängiges Verfahren der multivariaten Statistik, durch das sich optimale Kombinationen von gewichteten Merkmalen (allerdings meist ohne Berücksichtigung der Erkrankungshäufigkeiten für eine algorithmische Diagnostik errechnen lassen, ist die Fishersche Diskriminanzanalyse (vgl. Victor, 1973). Sind gewisse Voraussetzungen erfüllt, so erreicht man mit dieser Methode eine minimale Zahl von Fehlklassifikationen. Zu den Voraussetzungen gehört u.a., daß die Merkmale innerhalb der Gruppen multivariat normal verteilte, quantitative Größen sind. Die Methode kann außerdem nur dann angewendet werden, wenn für den Gruppenvergleich die Anzahl der Fälle größer als die Anzahl der Merkmale ist. Die erste Voraussetzung ist im psychiatrischen Bereich grundsätzlich nicht gegeben, und die zweite lässt sich aus äußeren Gründen oft nur schwer realisieren, da im allgemeinen für die psychiatrische Diagnosenstellung eine sehr große Anzahl von Merkmalen in Betracht zu ziehen ist. Man behilft sich deshalb damit, zunächst mittels Faktorenanalyse eine Datenreduktion vorzunehmen und gewinnt dadurch einen zusätzlichen Vorteil, da die Faktoren-Scores im allgemeinen eher die für eine Anwendung der Diskriminanzanalyse geforderten Verteileigenschaften besitzen als die Ausgangsdaten (etwa anamnestische Angaben oder Schätzwerte einzelner psychopathologischer Symptome). Baumann (1974) hat auf diese Weise anhand von psychopathologischen Befunddaten auf dem Umweg über faktoriell ermittelte Syndrom-Scores eine diskriminanzanalytische Trennung mehrerer klinisch-diagnostisch unterschiedener Patientengruppen mit den Kategorien Schizophrenien (kataton, paranoide und schizoaffektive Formen) und affektive Psychosen vorgenommen. Trotz der großen Anzahl von insgesamt 351 ausgewerteten Fällen wurde allerdings das Ergebnis nicht durch eine Kreuzvalidierung abgesichert.

Wir möchten nun im folgenden ein multivariates Verfahren der algorithmischen Diagnostik vorstellen, das von seinem Ansatz her eine im Verhältnis zum Umfang der verglichenen Gruppen große Anzahl von Merkmalen berücksichtigen kann und darüber hinaus auf die Verwendung von Alternativmerkmalen zugeschnitten ist. In der hier erprobten Version des Verfahrens werden jeweils zwei Gruppen miteinander verglichen. In diesem Sinne handelt es sich um ein speziell differential-diagnostisches Verfahren, dessen diagnostische Anwendung eine Vorentscheidung darüber voraussetzt, welche beiden Diagnosen für einen Fall überhaupt in Frage kommen. Allerdings können sukzessive Vergleiche von Diagnosenpaaren vorgenommen werden; doch liefert in diesem Falle das Verfahren nicht notwendigerweise eindeutige Ergebnisse, wie dies in einem hierarchisch gegliederten Entscheidungsverfahren, das von komplexeren zu einfacheren Kategorien fortschreitet, der Fall ist.

Da es sich bei unserer Untersuchung um eine Methodenstudie handelt, bei der es um die Erprobung einer 1973 von Holley angegebenen Methode in der algorithmischen psychiatrischen Diagnostik geht, begnügen wir uns — ebenso wie Baumann in seiner diskriminanzanalytischen Untersuchung — mit der Verwendung von Daten des psychopathologischen Befundes, beziehen also keine somatischen und keine anamnestischen Daten in die Analyse ein. Auch lassen wir außer acht, daß die Diagnose kein vom psychopathologischen Befund unabhängiges Klassifikationskriterium darstellt. Wir können das in Kauf nehmen, weil es sich ja lediglich um den Versuch einer Objektivierung der Diagnostik und nicht um eine Prüfung der Validität von Diagnosen handelt.

## B. Aufgabenstellung

Es ist das Ziel der vorliegenden Untersuchung, die diskriminanzanalytische Anwendung der *Q*-Faktorenanalyse (vgl. Überla, 1968) unter Verwendung des *G*-Index (Holley u. Guilford, 1964), die im folgenden *G*-Analyse genannt werden soll, zur Klassifikation von psychiatrischen Patienten zu überprüfen. Das Konzept der *G*-Analyse ermöglicht im Grunde eine genuine Gruppierung der Klassifikationsobjekte (Holley u. Fröbärlj, 1967), ähnlich wie es die Methoden der Clusteranalyse (vgl. Bock, 1970) erlauben. Wir wollen hier die *Q*-Analyse in einer teilweise kriteriumsorientierten Form verwenden, wie sie von Holley (1973) vorgeschlagen wurde. Ziel der Untersuchung ist es, Patienten jeweils einer von zwei vorgegebenen Gruppen richtig zuzuordnen. Als Kriterium für die Gruppenzugehörigkeit soll dabei die klinische Diagnose nach der internationalen Klassifikation der WHO (Degkwitz *et al.*, 1973) dienen. Dabei werden für jeweils zwei miteinander zu vergleichende Gruppen in den Analysestichproben die Items ausgewählt, die einzeln am besten zwischen diesen Gruppen differenzieren. Anhand der Analysestichproben wird also mit Hilfe von zwei psychiatrischen Schätzskalen ein diskriminanzanalytisches Verfahren zur differentialdiagnostischen Zuordnung konstruiert. In einem zweiten Schritt, der Kreuzvalidierung, wird die Güte dieses Zuordnungsverfahrens an einer unabhängigen Patientengruppe, der Validierungsstichprobe, überprüft.

## C. Methode

### 1. Datenerhebung

Das in der vorliegenden Arbeit verwendete Datenmaterial stammt aus der Erfassung des psychopathologischen Befundes von 454 Patienten (organische und endogene Psychosen, Neurosen und Persönlichkeitsstörungen), die in den Jahren 1969 und 1970 an der Psychiatrischen Abteilung des Max-Planck-Instituts für Psychiatrie in München und an anderen psychiatrischen Krankenhäusern in und um München im Rahmen eines Projektes zur Überprüfung von Schätzskalen untersucht wurden.

Von diesen Patienten wurden die folgenden Diagnosengruppen nach dem Schlüssel der ICD (International Classification of Diseases, Psychiatrischer Teil, 8. Rev., deutsche Ausgabe von Degkwitz *et al.*, 1973) ausgewählt, die für unsere Problemstellung eine genügend große Anzahl von Fällen umfaßten:

- 45 Fälle von Schizophrenie, paranoide Form (ICD-Nr. 295.3),
- 47 Fälle von Schizophrenie, nicht näher bezeichnete Form (ICD-Nr. 295.9),
- 44 Fälle von Depression im Rahmen einer affektiven Psychose (ICD-Nr. 296.2),
- 53 Fälle von depressiver Neurose (ICD-Nr. 300.4).

Zur Erfassung des psychopathologischen Befundes wurden zwei psychiatrische Schätzskalen verwendet, die Inpatient Multidimensional Psychiatric Scale (IMPS) von Lorr, Klett, McNair u. Lasky (Lorr *et al.*, 1962; Lorr *et al.*, 1963; Lorr u. Klett, 1967) und eine Skala aus dem Dokumentationssystem der deutsch-schweizerisch-österreichischen „Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie“ (AMP) (Angst *et al.*, 1967; Angst *et al.*, 1969; Scharfetter, 1972). Von dieser AMP-Skala verwendeten wir den Bogen für den psychopathologischen Befund mit 123 Items, ergänzt um 44 Items des Bogens für den somatischen Befund, die sich überwiegend auf vegetative, körperlich-funktionelle Beschwerden beziehen.

Die von uns verwendete Form der IMPS (deutsche Übersetzung: Flegel u. Mäder, unveröffentlicht) besteht aus 90 Einzel-Items zur Erfassung psychopathologischer Symptome, die in der Umgangssprache formuliert sind. Jedes Item muß vom Untersucher im Anschluß an ein freies Interview entsprechend einer neunfachen Abstufung (zwischen „fehlend“ = 0

und „extrem stark ausgeprägt“ (= 8) beurteilt werden. 75 dieser Items, nämlich die der ursprünglichen Version der Skala, ließen sich auf Grund von in den U.S.A. vorgenommenen Faktorenanalysen auf 10 Dimensionen aufteilen (Lorr u. Klett, 1966; Cohen, 1966), die sich fast in gleicher Form auch durch Untersuchungen im deutschen Sprachraum verifizieren ließen (Behrends et al., 1971; Jacobi, 1974; Bender, 1974; Mombour et al., 1973).

Die Items der AMP-Skala sind in psychiatrischer Fachterminologie formuliert. Sie müssen ebenfalls im Anschluß an ein freies Interview mit der Abstufung „nicht vorhanden“ (= 0), „leicht“ (= 1), „mittel“ (= 2) oder „schwer“ (= 3) beurteilt werden. Hinzu kommen die beiden Beurteilungskategorien „nicht näher beurteilbar“ und „fraglich vorhanden“, die von uns in der Auswertung mit „nicht vorhanden“ in einer Kategorie (= 0) zusammengefaßt wurden.

Die Faktorisierung der Items der AMP-Skala erbrachte für die oben genannte Stichprobe von 454 Patienten neun Faktoren, von denen sieben verschiedenen IMPS-Dimensionen zugeordnet werden konnten (Mombour et al., 1973). Für eine Stichprobe von 351 Patienten mit nur endogenen Psychosen ergaben sich ebenfalls neun Faktoren (Baumann, 1974), die überwiegend den von uns gefundenen ähnlich waren, wobei allerdings die bei unterschiedlichen Stichproben allgemein vorkommende Faktorenstrukturverschiebung in Erscheinung traten.

## 2. Darstellung der verwendeten Form der G-Analyse

a) *Itemselektion.* Bevor man diskriminanzanalytische Methoden anwenden kann, ist es im allgemeinen notwendig, die für die Trennung der Klassifikationsgruppen nützliche Information aus der insgesamt zur Verfügung stehenden Information zu extrahieren. Bei der linearen Diskriminanzanalyse kann man die Variablenzahl drastisch verringern, indem man das Verfahren der schrittweisen Diskriminanzanalyse so früh abbricht, daß nur die trennschärfsten Variablen verwendet werden.

Für die diskriminanzanalytische Verwendung der G-Analyse ist eine ähnlich starke Reduktion der Itemzahl nicht angebracht, doch ist auch bei ihr eine Einengung des Itempools auf möglichst trennscharfe Items vorteilhaft.

Verwendet man alle erfaßten Items, so befindet sich unter ihnen auch eine Reihe, die nichts zur Trennung der Gruppen beiträgt und die daher ausgesondert werden kann. Um die Unabhängigkeit der Validierungsstichprobe zu erhalten, dürfen diese Items nur auf Grund der jeweiligen Analysestichprobe bestimmt werden. Die Items werden so zwar stichproben-spezifisch auf die Analysegruppe hin ausgewählt, allerdings mit der Vermutung, daß dadurch auch für die Validierungsgruppe irrelevante Information vermieden wird.

Als einfaches Maß zur Selektion der Items haben wir die Differenz der Mittelwerte in den beiden Teilgruppen ausgewählt, die von Holley u. Risberg (1972) für dichotome Merkmale als *D*-Estimate vorgeschlagen wurde. Dieses Maß ist numerisch der  $\varphi$ -Korrelation zwischen dem Item und der Variablen „Gruppenzugehörigkeit“ gleich, wenn die Teilgruppen und die Alternativen des dichotomen Merkmals gleich häufig vorkommen. Wenn man das Zuordnungsverfahren optimieren will, wird man im konkreten Anwendungsfall versuchen, die Grenze für die *D*-Estimates so zu bestimmen, daß durch das diskriminanzanalytische Verfahren insgesamt die Zahl der richtigen Zuordnungen für die Analysestichprobe maximiert wird. Im vorliegenden Falle wurde willkürlich festgelegt, die Items nicht in die Analyse einzubeziehen, deren *D*-Estimate zwischen -0,10 und +0,10 lag. Die Entscheidung für  $|D| \geq 0,10$  stellt einen Kompromiß dar; einerseits besteht der Wunsch nach gut trennenden Items, andererseits darf die Zahl der Items im Vergleich zu der der Fälle bei der *Q*-Analyse nicht zu klein sein. Die trennschärfsten Items sind in Tab. 3 aufgeführt.

b) *Der G-Index als Korrelationskoeffizient.* Die *Q*-Faktorenanalyse baut auf den Interkorrelationen der Personen auf. Während bei der Interkorrelation von Variablen über Personen die Orientierung der zugrundeliegenden Skalen nur das Vorzeichen, nicht aber die Höhe des Korrelationskoeffizienten beeinflußt, sind Korrelationen zwischen Personen über Variablen wesentlich abhängig von der Polung der Variablen (vgl. Baumann, 1971). Für die *Q*-Faktorenanalyse ist es deshalb wünschenswert, einen Koeffizienten heranzuziehen, der von der Polung der Variablen unabhängig ist. Geht man von alternativen Daten aus, mit denen man im klinischen Bereich die vorhandene Information in vielen Fällen ausschöpft, erreicht man dies bequem dadurch, daß man zu jeder Variablen  $X_j$  ihre umgepolte Variante  $\bar{X}_j$  auf-

nimmt. Demonstrieren wir dies an der Vierfeldertafel für die Korrelation zwischen Person  $P_1$  und Person  $P_2$ . Für den einfachen Variablenatz erhalte man Tab. 1 mit der Produkt-Moment-Korrelation

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Tabelle 1. Vierfeldertafel für den  $\varphi$ -Koeffizienten (einfacher Variablenatz)

		$P_1$		
		+	-	
$P_2$	+	$a$	$b$	$a + b$
	-	$c$	$d$	$c + d$
		$a + c$	$b + d$	$n$

Tabelle 2. Vierfeldertafel für den  $G$ -Index (erweiterter Variablenatz)

		$P_1$		
		+	-	
$P_2$	+	$a + d$	$b + c$	$n$
	-	$c + b$	$d + a$	$n$
		$n$	$n$	$2n$

Bei der Erweiterung des Variablenatzes geht für jede Kombination ++ eine Kombination -- in die Häufigkeitstabelle ein und für jedes +- ein -+ usw. Man erhält die Tab. 2 und als Produkt-Moment-Korrelation

$$G = \frac{(a+d)(d+a) - (b+c)(c+b)}{\sqrt{n \cdot n \cdot n \cdot n}},$$

die unter der Verwendung des Anteils übereinstimmender Antworten  $P_c = \frac{a+d}{n}$  übergeht in  $G = 2P_c - 1$ .

Dieser  $G$ -Index, den wir der  $Q$ -Faktorenanalyse zugrunde gelegt haben, ist einerseits als Produkt-Moment-Korrelation und andererseits als Anteil übereinstimmender Antworten interpretierbar. Der  $G$ -Index wurde als Korrelationsmaß zwischen Personen zur Verwendung in der  $Q$ -Faktorenanalyse von Holley u. Guilford (1964) vorgeschlagen; Signifikanztests wurden von Lienert (1972) entwickelt; seine Eigenschaften und die Probleme bei der Übertragung auf  $R$ -Faktorenanalysen werden von Levy (1966, 1967) und von Holley (1966, 1967) diskutiert.

c) *Faktorisierung der Analysestichprobe nach der Q-Methode.* Zur Errechnung der  $G$ -Indices wurden alle Items derart dichotomisiert, daß die Alternativen weitgehend gleich häufig auftraten. Für die IMPS wurden die Alternativen 1 bis 8 zu + 1, die Alternative 0 zu - 1 transformiert, für die AMP wurden die Antworten „leicht“, „mittel“ und „schwer ausgeprägt“ durch + 1, die Antworten „nicht vorhanden“, „nicht beurteilbar“, „fraglich vorhanden“ durch - 1 dargestellt.

Entsprechend unserer Aufgabenstellung, bei der für die Patienten zweier jeweils vorgegebener Diagnosengruppen die richtige Zuordnung zu einer der beiden Diagnosen angestrebt wurde, führten wir eine  $Q$ -Faktorenanalyse für jedes Diagnosenpaar getrennt durch. Für jedes Diagnosenpaar skalierten wir also die Personen auf zwei Dimensionen, den varianzstärksten Faktoren der  $Q$ -Analyse (Faktorisierung nach dem Hauptkomponentenmodell, vgl. Überla, 1968; Gaensslen u. Schubö, 1973), und wählten dann eine Gerade durch den Ursprung als Trennlinie zwischen den beiden Gruppen. Als Trennlinie kann ohne weiteres eine Winkelhalbierende verwendet werden, wenn man die beiden Faktoren nach dem Varimax-Kriterium von Kaiser orthogonal rotiert hat.

Das Zuordnungsverfahren sei am Beispiel der Gruppen „Schizophrenie, nicht näher bezeichnete Form“ und „Depression im Rahmen einer affektiven Psychose“ erläutert (s. Abb. 1). In diesem Beispiel entspricht der Faktor I der Gruppe „Schizophrenie, nicht näher bezeichnete Form“, der Faktor II der Gruppe „Depression im Rahmen einer affektiven Psychose“. Unser Zuordnungsverfahren besteht darin, daß eine Person dann der Schizophreniegruppe zugeordnet wird, wenn ihre Ladung auf Faktor I größer als die auf Faktor II ist, anderenfalls wird sie der bezeichneten Depressionsgruppe zugeordnet.

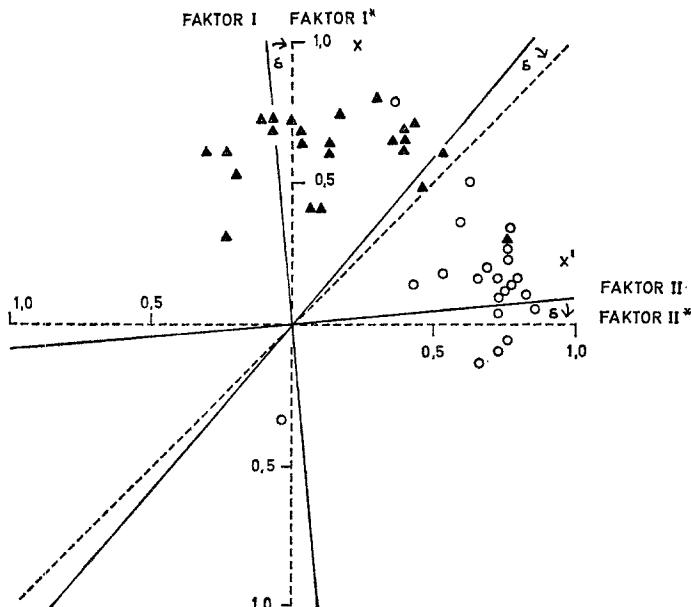


Abb. 1. Lage der Personen der Analysestichproben im Achsensystem der Faktoren: ▲ Schizophrenie, nicht näher bezeichnete Form; ○ Depression im Rahmen einer affektiven Psychose. Die Stabilisatoren sind mit  $x$  bzw.  $x'$  gekennzeichnet. Die gestrichelten Linien bezeichnen das um den Winkel  $\delta$  nachgezeichnete Achsensystem und die mitrotierte Trennlinie

*d) Stabilisatoren und Nachrotation.* Neben der Gewichtung der Variablen durch die Itemselektion führten wir auch eine Akzentuierung der Personengruppen durch. Dies geschah durch Einführung artifizieller typischer Personen („Stabilisatoren“) in die Analysengruppe. Nach einem Vorschlag von Holley<sup>1</sup> werden zu jeder der beiden Teilgruppen der Analysengruppe die Mittelwerte in allen Items berechnet. Die für eine Teilgruppe typische Person besteht aus den Mittelwerten der Items für die Teilgruppe. Um das Gruppenmittel im Verhältnis zu den realen Personen stärker zu gewichten, wird zusätzlich zu diesen die typische Person fünfmal aufgenommen. Die Lage der Stabilisatoren für unser Beispiel ist in Abb. 1 eingezzeichnet. In dieser Abbildung wird deutlich, daß die Stabilisatoren nicht symmetrisch zum Achsensystem Faktor I/Faktor II liegen müssen. Durch Nachrotation des Koordinatenystems um den Winkel  $\delta$  wird eine symmetrische Lage der Stabilisatoren im neuen Achsensystem Faktor I\*/Faktor II\* erreicht (gestrichelte Linien in Abb. 1). Die Zahl der nach dem Kriterium der jeweiligen Diagnose für die Analysestichprobe richtig und falsch zugeordneten Personen ist aus Tab. 3 ersichtlich.

*e) Übertragung auf die Validierungsstichprobe.* Bei allen bisher beschriebenen Klassifikationen und der Entscheidung für die Stabilisatoren, Itemselektion und Nachrotation wurde von den Daten der Validierungsgruppe kein Gebrauch gemacht. Es geht nunmehr darum, das an der Analysengruppe entwickelte Klassifikationsmodell auf die Validierungsgruppe zu übertragen und hier seine Brauchbarkeit zu prüfen.

Zur Zuordnung benötigen wir, wie dargestellt, die Ladungen der Personen in den beiden Faktoren. Im Falle orthogonaler Faktoren sind die Ladungen als Korrelationen zwischen den Personen und den Faktoren über die Items bestimmt (vgl. Gaensslen u. Schubö, 1973; Überla, 1968). Da die Personen der Validierungsgruppe nicht in die Faktorisierung der Analysengruppe einbezogen werden können, liegen für sie keine Ladungen vor. Es lassen sich aber die Korrelationen zwischen diesen zusätzlichen Personen und den Faktoren der Analysengruppe be-

<sup>1</sup> Mündliche Mitteilung.

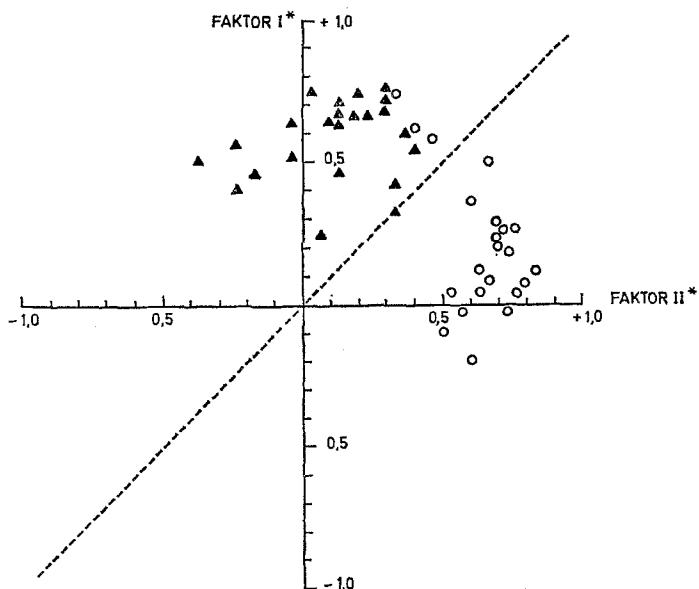


Abb. 2. Lage der Personen der Validierungsstichproben im Achsensystem der nachrotierten Faktoren: ▲ Schizophrenie, nicht näher bezeichnete Form; ○ Depression im Rahmen einer affektiven Psychose

rechnen, indem man jeweils die Item-Werte einer Person mit den Faktorenwerten der Items im Faktor korreliert. Dadurch erhält man eine Schätzung der diese Personen kennzeichnenden Faktorenladungen, von denen angenommen wird, daß sie sich bei gemeinsamer Faktorisierung von Analyse- und Validierungsgruppen ergeben hätten. Die Personen der Validierungsgruppe werden dem Faktor zugeordnet, mit dem sie am höchsten korrelieren. Abb. 2 zeigt das Ergebnis der Klassifikation wieder am Beispiel der Gruppen „Schizophrenie, nicht näher bezeichnete Form“ und „Depression im Rahmen einer affektiven Psychose“ für die Validierungsgruppe.

#### D. Ergebnisse

Für die klinische Interpretation der Klassifikationsergebnisse sind die Faktorenwerte des Items wichtig, die zur Schätzung der Faktorenladungen in der Validierungsgruppe verwendet wurden. Sie bedeuten nämlich die Gewichte, mit denen einzelne Items bei der Gruppentrennung wirksam werden. Bei der Itemselektion mit Hilfe der *D*-Estimates ergaben sich jeweils etwa 100 Items. Für die trennschärfsten sind in Tab. 3 die Faktorenwerte aufgeführt.

Die Häufigkeiten der richtigen und falschen Zuordnungen für alle Gruppenpaare der Validierungsstichprobe sind in Tabelle 4 den Ergebnissen für die Analysestichprobe gegenübergestellt. Die Übereinstimmung der paarweisen Zuordnung mit der psychiatrischen Diagnose liegt für die Validierungsgruppen im Mittel bei 83 %. Die Zahl der richtigen Zuordnungen wurde auf ihre Abweichung von einer rein zufälligen Klassifikation für jedes der sechs Klassifikationsprobleme mit dem Binomialtest überprüft. Alle Zuordnungen außer der für die Trennung von „Depression im Rahmen einer affektiven Psychose“ und „Depressive Neurose“ weichen hoch signifikant ( $P < 0.00001$ ) von der Zufallszuordnung ab.

Tabelle 3. Die für die Trennung der Diagnosegruppen bedeutsamsten Items aus IMPS und AMP-Skala und ihre jeweiligen Faktorwerte

Nr.	Item	Faktorwert für 295.3 / 295.5	Faktorwert für 295.3 / 296.2	Faktorwert für 295.3 / 300.4	Faktorwert für 295.5 / 296.2	Faktorwert für 295.5 / 300.4	Faktorwert für 296.2 / 300.4
<b>IMPS<sup>a</sup></b>							
2	Vorbeireden	.0	-1.4	.0	-1.4	.0	-1.4
5	äußerst feindselige Gefühle	.3	-1.1	.3	-1.1	.5	-1.3
8	gleichgütig gegenüber eigener Zukunft	2.4	-.9	1.3	-.3	.5	-.1
10	Gespanntheit	-.4	-1.1				
14	Selbstbeschuldigungen	.1	-1.1				
18	feindselige Haltung	1.4	-1.1				
19	Mangel an Erinnerungsvermögen	.5	-1.3				
21	definierte Ängste	.5	-1.6	.9			
22	stockt beim Sprechen						
23	Apathie						
24	unbestimmte Ängste						
27	Minderwertigkeitsgefühle						
28	beschuldigt andere						
29	deprimiert						
31	Schuldgefühl						
36	mißtrauisch						
38	antwortet nicht oder einsilbig						
39	Mangel an Einsicht						
44	wahnhaft						
45	hört Stimmen						
46	läppisch						
59	Beziehungsideen	1.2	-1.2	.9	-1.5	.8	-1.9
60	Verfolgungsideen						
61	hält sich für Opfer einer Verschwörung						
76	spricht mit tiefem Gefühl						
78	hoffnungslos						
81	Libidoverlust						
82	Gewichtsverlust						
83	Appetitverlust						
84	Unfähigkeit etwas zu Wege zu bringen	.8	-1.3				
86	Schlafstörungen	1.0	-1.4	-.7	1.2	-.6	1.3
87	Energienangel						
90	Interessenverlust						

<sup>a</sup> abgekürzte Formulierung.

Tabelle 3 (Fortsetzung)

Nr. Item	Faktorwert für 295.3 / 295.5	Faktorwert für 295.3 / 296.2	Faktorwert für 295.3 / 300.4	Faktorwert für 295.5 / 296.2	Faktorwert für 295.5 / 300.4	Faktorwert für 296.2 / 300.4
<b>AMP-Skala<sup>a,b</sup></b>						
23 Vorbeireden	-2	-1.3	-2	-1.4		
24 inkohärent zerfahren	-3	-1.3	-3	-1.4		
34 Wahnstimmung	-1	-1.3	-1	-1.6		
36 Wahneinfall, Wahngedanken	-1	-1.3	-1	-1.6		
37 Wahndynamik	-4	-1.2	-3	-1.4		
43 Beziehungswahn	.8	-1.1	.8	-1.5	.6	-1.7
44 Beeinträchtigungs-/Verfolgungswahn			.3	-1.3	.2	-1.6
47 Illusionen	-1	-1.5	-5	-1.1	-4	-1.3
48 Stimmenhören						-1.7
60 sonstige Fremdbeeinflussungserlebnisse	.1	-1.0				-2
62 ratlos						-1.5
63 gefühlsverarmt/affektarm				.3	-1.0	
65 Störung der Vitalgefühle				-1.7	.4	
66 deprimiert/traurig	-1.3	1.6	-1.6	-1.0	-1.6	.4
67 hoffnungslos/verzweifelt	-1.9	1.4	-1.6	-1.7	1.2	-1.7
68 ängstlich				-1.2	.6	-1.1
71 müßtranisch/feindselig						.0
72 gespannt	1.8	-1.1	.7	-1.3	.2	-1.5
73 innerlich unruhig	1.0	-1.3	.8	-1.0		
74 klagesam/jammerig						
77 Insuffizienzgefühl						
79 Schuldgefühl						
82 affektiv inadäquat						
85 affektstarr						
99 abends besser						
106 Mangel an Krankheitsgefühl	-2.2	.8	-2	-1.3		
107 Mangel an Krankheitseinsicht	.6	-1.5	.4	-1.4		
109 Suizidtendenzen						

1155 Sexualität vermindert		-2.1	.9	-1.1	.8	.8	-1.3
1244 Einschlafstörungen	.9	-1.4	-.5	1.1	-1.4	.8	
1255 Durchschlafstörungen			-1.7	.6	-1.6	1.4	-1.1
1266 Verkürzung der Schlafdauer					-1.6	.4	
1288 Müdigkeit					-1.4	-1.7	.5
1422 Obstipation	.0	-1.3			-1.4	-1.7	.7
1511 Herzbeunsten					-1.3	-1.1	
1522 Schwindel					-1.6	-1.6	.3
					-1.7	.4	

„b“ fortlaufende Numerierung der Reihenfolge der Skalen-Items.

Tabelle 4. Übereinstimmung der formalen Klassifikation mit Hilfe der G-Analyse mit der vom Psychiater gestellten Diagnose

Gruppenpaare psychiatrische Diagnose	ICD-Nr.	Analysestichprobe			Validierungsstichprobe		
		Gruppen- größe	richtig zugeordnet	falsch zugeordnet	Gruppen- größe	richtig zugeordnet	falsch zugeordnet
Schizophrenie: paranoide Form	295.3	23	14	9	22	19	3
Schizophrenie: nicht näher bezeichnete Form	295.9	24	19	70%	23	19	84%
Schizophrenie: paranoide Form	295.3	23	17	6	22	19	3
Depression im Rahmen einer affektiven Psychose	296.2	22	20	82%	2	22	21
Schizophrenie: paranoide Form	295.3	23	17	6	22	19	3
Depressive Neurose	300.4	27	21	76%	6	26	24
Schizophrenie: nicht näher bezeichnete Form	295.9	24	23	1	23	22	1
Depression im Rahmen einer affektiven Psychose	296.2	22	21	96%	1	22	19
Schizophrenie: nicht näher bezeichnete Form	295.9	24	23	1	23	22	1
Depressive Neurose	300.4	27	21	86%	6	26	19
Depression im Rahmen einer affektiven Psychose	296.2	22	18	71%	4	22	15
Depressive Neurose	300.4	27	17	10	10	26	13
Mittelwerte für alle Gruppenpaare					80%		83%

Aus Tabelle 4 ist ersichtlich, daß das formale Klassifikationsverfahren in der beschriebenen differentialdiagnostischen Anwendung recht valide ist, wenn man berücksichtigt, daß es nicht alle Informationen benutzen kann, die dem Psychiater zur Verfügung stehen. Dabei muß wohl als Zufall gewertet werden, daß die Ergebnisse in der Validierungsstichprobe im Mittel etwas besser ausfallen als in der Analysestichprobe.

### E. Diskussion

Das Ergebnis der Analyse erscheint klinisch sehr plausibel: Diejenigen Gruppen, die sich definitionsgemäß durch verschiedenartige Psychopathologie voneinander abheben, wie schizophrene Psychosen und Depressionen (gleich welcher Genese) sowie paranoide und nicht-paranoide Formen schizophrener Psychosen werden in den Validierungsstichproben durch das algorithmische Vorgehen hochsignifikant voneinander getrennt. Die Diskrepanzen dürften etwa in der Größenordnung liegen, wie sie beim Vergleich der von verschiedenen Klinikern gleicher nosologischer Orientierung gestellten Diagnosen zu erwarten sind. Der Prozentsatz richtiger Zuordnungen würde sich durch die Berücksichtigung anamnestischer Daten kaum noch wesentlich verbessern lassen. Allerdings ist zu berücksichtigen, daß vor dem hier vorgenommenen Vergleich schon wesentliche diagnostische Vorentscheidungen gefallen waren, nämlich der Ausschluß andersartiger als der jeweils in Frage kommenden Diagnosen.

Beim diagnostischen Vergleich von Depressionsformen verschiedener Genesen versagt der durch den Vergleich der Analysestichprobe gewonnene Algorithmus. Dies dürfte aber nicht auf eine unzureichende Ausschöpfung differentialdiagnostisch relevanter Information durch das verwendete Analyseverfahren zu beziehen sein, sondern auf das im Datenmaterial selber begründete Informationsdefizit: Der psychopathologische Befund allein genügt wohl zur Diagnose eines depressiven Syndroms, nicht aber zur Unterscheidung ätiologisch differenzierbarer Depressionsformen. Dafür sind anamnestische Daten unentbehrlich. Dieser Mangel kann selbstverständlich durch kein Verfahren der Datenanalyse ausgeglichen werden. Da die Differentialdiagnose verschiedener Depressionsformen von großem wissenschaftlichen wie auch praktisch-klinischem Interesse ist (u. a. für die Prognosenstellung und für das therapeutische Vorgehen), erscheint eine Anwendung des hier dargestellten Analyseverfahrens auf ein um entsprechende anamnestische Angaben erweitertes Datenmaterial von Patienten mit endogenen bzw. neurotischen Depressionen besonders wichtig. Es sollte allerdings auch geprüft werden, ob das Holleysche Analyseverfahren unter den hier geschilderten Voraussetzungen bessere Ergebnisse liefert als andere Vorgehensweisen einer algorithmischen Diagnostik.

Als weitere Entwicklungsmöglichkeit der hier vorgestellten Methodik bietet sich der Gruppenvergleich über mehr als zwei Faktoren an. Dies erscheint relevant besonders bei der Ausweitung des Verfahrens auf die simultane Unterscheidung von mehr als zwei Gruppen.

### Literatur

- Angst, J., Battegay, R., Bente, D., Berner, P., Broeren, W., Cornu, F., Dick, P., Engelmeier, M. P., Heimann, H., Heinrich, K., Helmchen, H., Hippius, H., Pöldinger, W., Schmidlin, P., Schmitt, W., Weis, P.: Über das gemeinsame Vorgehen einer deutschen

- und schweizerischen Arbeitsgruppe auf dem Gebiet der psychiatrischen Dokumentation. Schweiz. Arch. Neurol. Neurochir. Psychiat. **100**, 207—211 (1967)
- Angst, J., Battegay, R., Bente, D., Berner, P., Broeren, W., Cornu, F., Dick, P., Engelmeier, M. P., Heimann, H., Heinrich, K., Helmchen, H., Hippius, H., Pöldinger, W., Schmidlin, P., Schmitt, W., Weis, P.: Das Dokumentations-System der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMP). Arzneimittel-Forsch. **19**, 399—405 (1969)
- Baumann, U.: Psychologische Taxonometrie. Bern: Huber 1971
- Baumann, U.: Diagnostische Differenzierungsfähigkeit von Psychopathologie-Skalen. Arch. Psychiat. Nervenkr. **219**, 89—103 (1974)
- Bégon, F., Trémolières, R., Sultan, C.: Présentation et illustration d'un modèle logique du diagnostic à propos d'un exemple hématologique. Journées d'informatiques médical, Bd. 2, S. 33—52. Toulouse 1972
- Behrends, K., Flegel, H., Helmchen, H., Hippius, H., Höffken, K. D., Schacht, L., Schulte, P. W.: Quantifizierung psychotischer Symptome unter transkulturellen Aspekten. Soc. Psychiat. **6**, 66—72 (1971)
- Bender, W.: Studie zur Reliabilität und differentiellen Validität der Lorr-Skala (IMPS). Dissertation, Universität Hamburg 1974
- Blomer, R. J., Victor, N., Schimetzek, H.: A dialogue-orientated model for computer-aided diagnosis. Journées d'informatiques médical, Bd. 2, S. 69—79. Toulouse 1972
- Bock, H. H.: Automatische Klassifikation. In: Statistische Methoden II, E. Walter (Hrsg.), S. 36—80. Berlin-Heidelberg-New York: Springer 1970
- Cohen, J.: The impact of multivariate research in clinical psychology. In: Handbook of multivariate experimental psychology. R. B. Cattell (Hrsg.). Chicago: Rand McNally 1966
- Degkwitz, R., Helmchen, H., Mombour, W. (Hrsg.): Diagnosenschlüssel und Glossar psychiatrischer Krankheiten, 3. Aufl. Berlin-Heidelberg-New York: Springer 1973
- Gaenslen, H., Schubö, W.: Einfache und komplexe statistische Analyse. München: Reinhardt 1973
- Gerster, F., Dirlich, G.: Ein Programm zur Entscheidungshilfe in der Psychiatrie. IBM Nachrichten **25**, 44—48 (1975)
- Gorry, G. A.: A system for computer-aided diagnosis. Project MAC, technical report MAC-44. Cambridge, Mass.: M. I. T. 1967
- Holley, J. W.: A reply to Philip Levy—in defense of the G-index. Scand. J. Psychol. **7**, 244—246 (1966)
- Holley, J. W.: Philip Levy and the G-index—a final reply. Scand. J. Psychol. **8**, 250 (1967)
- Holley, J. W.: Rorschach Analysis. In: New approaches in psychological measurement, P. Kline (Hrsg.), S. 119—155. London: Wiley 1973
- Holley, J. W., Fröbärg, G.: A demonstration study on the validity of the Rorschach test. Int. Rev. appl. Psychol. **16**, 30—51 (1967)
- Holley, J. W., Guilford, J. P.: A note on the G-index of agreement. Educ. Psychol. Measmt. **14**, 749—753 (1964)
- Holley, J. W., Risberg, J.: On the D estimate of discriminatory effectiveness. Psychol. Res. Bull. Lund Univ. **12** No. 12 (1972)
- Horrocks, J. C., McCann, A. P., Staniland, J. R., Leaper, D. J., de Dombal, F. T.: Computer-aided diagnosis: Description of an adaptable system and operational experience with 2034 cases. Brit. med. J. **1972 II**, 5—9
- Jacobi, P.: Untersuchungen zur Faktorenstruktur, Zuverlässigkeit und Gültigkeit einer deutschen Bearbeitung der IMPS nach Lorr. Dissertation, Universität Hamburg 1974
- Jesdinsky, H. J.: Deterministische Zuordnungsverfahren. In: Computerunterstützte ärztliche Diagnostik, H.-J. Lange u. G. Wagner (Hrsg.), S. 237—244. Stuttgart-New York: Schattauer 1973
- Lange, H. J.: Algorithmische Diagnostik. Münch. med. Wschr. **113**, 577—580 (1971)
- Ledley, R. S.: Use of computers in biology and medicine. New York: McGraw-Hill 1965
- Levy, P.: Properties of the Holley-Guilford "G-index of agreement" in R and Q factor analysis. Scand. J. Psychol. **7**, 239—243 (1966)
- Levy, P.: A reply to Jasper Holley's defense of the G-index. Scand. J. Psychol. **8**, 38 (1967)

- Lienert, G. A.: Note on tests concerning the index of agreement. *Educ. Psychol. Measmt.* **32**, 281–288 (1972)
- Lorr, M., Klett, C. J.: Manual for the inpatient multidimensional psychiatric scale (IMPS); revised. Palo Alto/Calif.: Consulting Psychologists Press 1967
- Lorr, M., Klett, C. J., McNair, D. M.: *Syndromes of psychosis*. Oxford-London-New York-Paris: Pergamon Press 1963
- Lorr, M., Klett, C. J., McNair, D. M., Lasky, J. J.: Inpatient multidimensional psychiatric scale (IMPS). Washington, D.C.: Veterans Administration 1962
- Mombour, W.: Verfahren zur Standardisierung des psychopathologischen Befundes, Teil I/II. *Psychiat. Clin.* **5**, 73–120, 137–157 (1972)
- Mombour, W.: Syndrome bei psychiatrischen Erkrankungen. Eine vergleichende Untersuchung mit Hilfe von zwei Schätzskalen für den psychopathologischen Befund (IMPS und AMP-Skala). *Arch. Psychiat. Nervenkr.* **219**, 331–350 (1974)
- Mombour, W., Gammel, G., v. Zerssen, D., Heyse, H.: Die Objektivierung psychiatrischer Syndrome durch multifaktorielle Analyse des psychopathologischen Befundes. *Nervenarzt* **44**, 352–358 (1973)
- Scharfetter, Ch.: Das AMP-System, 2. Aufl. Berlin-Heidelberg-New York: Springer 1972
- 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)
- Schmid, W., Castell, R., Mombour, W., Mittelsten Scheid, D., v. Zerssen, D.: Die diagnostische Übereinstimmung zwischen Klinikern und dem DIAL-Programm. *Arch. Psychiat. Nervenkr.* **218**, 339–351 (1974)
- Überla, K.: Faktoren-Analyse. Berlin-Heidelberg-New York: Springer 1968
- Victor, N.: Probabilistische Zuordnungsverfahren. In: Computerunterstützte ärztliche Diagnostik, H.-J. Lange u. G. Wagner (Hrsg.), S. 245–251. Stuttgart-New York: Schattauer 1973
- Wing, J. K., Cooper, J. E., Sartorius, N.: Measurement and classification of psychiatric symptoms. London: Cambridge University Press 1974
- Zerssen, D. v.: Diagnose. In: Lexikon der Psychiatrie, C. Müller (Hrsg.), S. 135–139. Berlin-Heidelberg-New York: Springer 1973

W. Schubö, Dipl.-Phys.  
Institut für Psychologie der Universität  
D-8000 München 40  
Friedrichstr. 22  
Bundesrepublik Deutschland

Dr. U. Hentschel, Dipl.-Psych.  
Prof. Dr. med. D. v. Zerssen, Dipl.-Psych.  
Dr. med. W. Mombour  
Max-Planck-Institut für Psychiatrie  
D-8000 München 40  
Kraepelinstr. 10  
Bundesrepublik Deutschland