

**Sonderdrucke aus**  
**DIAGNOSTICA**

**Heft 3 / 1977**

# Zur Erfassung von Hirnschädigungen bei Kindern: Nichtlineare Entscheidungsregeln auf der Basis von Veränderungsmessungen und des Jackknife\*

Claus Möbus & Roland Wallasch

Psychologisches Institut der Universität Heidelberg

## 1. Einleitung

Die Notwendigkeit, ein in der Durchführung ökonomisches und in der Prognose exaktes Instrument zur Erfassung von Hirnschädigungen bei Kindern zu konstruieren, ist – auch in der modernen Verhaltensdiagnostik (Kanfer & Saslow, 1974; Schulte, 1974) – unbestritten. Dieses ist auf dem Hintergrund anscheinend weltweit steigender Zahlen von Kindern mit Hirnschäden zu sehen. Die Gründe hierfür sind vielfältig und nicht allein auf ein durch verbesserte Diagnosemethoden erzeugtes Artefakt reduzierbar. Sie reichen von den größeren Überlebenschancen von „Risikokindern“ bis zur ständig zunehmenden Verwicklung von Kindern in Verkehrsunfällen (Birch, 1964; Paine, 1965; Prüssing, 1967; Lempp, 1967; Mumenthaler, 1971; Scholtz, 1972; Klee, 1974).

Besonders im schulischen Bereich wirken sich Hirnschädigungen störend aus. So weisen eine Reihe von Autoren (Lynn, 1960; Lempp, 1964, 1971; Paine, 1965; Halstead & Rennick, 1966; Frierson & Barbe, 1967; Myklebust & Boshes, 1967; Regel, 1967; Heck, 1969; Johnson & Myklebust, 1971; Tarnopol, 1971; Keogh et al., 1968) auf den Bezug zu Lernschwierigkeiten, mangelnde Schulreife (Müller, 1967; Kornmann, 1971), Hyperkinese oder Hyperaktivität (Boder, 1965; Clements & Peters, 1965; Clements, 1966; Dallmeyer, 1969; Kuhlen, 1972; Werry, 1968, 1972), psychopathologische Auffälligkeiten (incl. der durch die Reaktion der Umwelt auf die Hirnschädigungen entstandene „sekundäre Neurotisierung“) hin (Göllnitz, 1961; Boder, 1965; Graham & Rutter, 1968; Bräutigam, 1969; Dallmeyer, 1969; Müller-Küppers, 1969; Specht, 1969; Schulte & Tölle, 1971; Lempp, 1964, 1967, 1971; Wing, 1973). Darüber hinaus scheinen auch Bezüge zu kindlichen Psychosen und zum kindlichen Autismus herstellbar zu sein (Bender, 1955; Birch, 1964; Schulte & Tölle, 1971; Werry & Quay, 1972).

Eine effiziente Frühdiagnostik der cerebralen Schädigung oder Dysfunktion ist daher wünschenswert zur Planung der psychotherapeutischen Intervention, wenn sie auf Prävention psychischer Deviation und auf psychohygienische Maßnahmen abzielt. Die Bedeutung psychologischer Testverfahren ist trotz ihrer z. T. jahrzehntelangen Anwendungsgeschichte nicht zuletzt wegen der an eine Individualdiagnostik zu stellenden strengen Anforderungen umstritten (vgl. a. Billingslea, 1963; Reitan, 1962; Haynes & Sells, 1963; Yates, 1954, 1966; Reitan & Heineman, 1968;

\* leicht veränderte Fassung des Institutsberichts Nr. 5 vom August 1976.

Adams, 1969; Duhm, 1969; Burgess et al., 1970; Zimet & Fishman, 1970; Cronbach, 1970; Kornmann, 1971; Kaspar & Schulman, 1972; Schlange et al., 1972; Werry, 1972). Die univariat erzielten Ergebnisse befriedigen insgesamt jedoch nicht.

Einen Ausweg schien der multivariate Ansatz zu bieten. Um die konkurrente Validität und damit die Gesamttrefferquote zu erhöhen, wurden Testbatterien erstellt und die Prognosen mit multiplen Entscheidungsregeln (multiple Regression, Diskriminanzanalyse) getroffen (Harper, 1950; Pichot & Perse, 1952; Wewetzer, 1959; Stein, 1961; Reitan, 1962; Haynes & Sells, 1963; v. Kerekjarto, 1963; Wheeler, 1961, 1964; Burgess et al., 1970; Kornmann, 1971). Dieses Vorgehen erscheint plausibel und der Mehrdimensionalität des Hirnschadenskonzeptes angemessen. Es hat sich aber gezeigt (z. B. Wallasch & Möbus, 1976), daß die mit multiplen Ansätzen erzielbaren Trefferquoten einer Kreuzvalidierung mit der one-held-out Technik von Tukey (1958) und Lachenbruch (1965) nicht standhielten und auf die Trefferquote des besten Einzelprädiktors (Background Interference Procedure (BIP) mit dem von Schlange im Göttinger Formreproduktions-test (GFT) benutzten Auswertungssystem) zurückfiel. Auch hier bestätigte sich die in der Diagnostik schmerzlich gewonnene Erkenntnis: je größer die Zahl der zu schätzenden Parameter (z. B. Regressions- oder Diskriminanzgewichte), desto instabiler die Ergebnisse und desto stärker die Schrumpfung der kreuzvalidierten Trefferquoten.

Ziel unserer Untersuchung ist es, einen Mittelweg zwischen uni- und multivariater Methodik zu finden, in die Entscheidungsregel möglichst wenig zu schätzende Parameter einzubringen, die bedingte Trefferwahrscheinlichkeit  $P[\text{Diagnose} = \text{Hirnorganiker (HO)} \mid \text{Test} = \text{HO}]$  zu maximieren und die bedingte Fehlerwahrscheinlichkeit  $P[\text{Test} = \text{HO} \mid \text{Diagnose} = \text{Normal (N)}]$  zu minimieren. Wir stellen uns hiermit in gewissem Gegensatz zu den meisten Testautoren, die den Akzent hauptsächlich auf die Maximierung der Gesamttrefferquote und damit des Validitätskoeffizienten richten. Nur bei perfekter Validität (100%ige Trefferquote) sind beide Kriterien identisch. Der Grund für unsere etwas anders geartete Zielsetzung liegt in der besonderen Problematik der Hirnschadensdiagnostik, in der eine irrtümliche Hirnschadensprognose bei Normalen strikt vermieden werden muß. Wie im Teil 3 gezeigt wird, ist dieses Ziel nicht durch eine einfache Verschiebung der cut-off Werte, sondern nur durch eine multiplikative Verknüpfung der Einzelprädiktoren erreichbar.

## 2. Rahmenbedingungen der Untersuchung: Pbn und Tests

Die Untersuchung basiert auf einer Gruppe hirngeschädigter und einer Gruppe normaler Kinder (à 48 Pbn), die bezüglich Alter und IQ (HAWIK) durch Paarbildung (statistische Zwillinge) parallelisiert wurden. Als hirngeschädigt wurden solche Kinder bezeichnet, die von Fachärzten aufgrund sorgfältiger<sup>3</sup> Untersuchungen als solche diagno-

<sup>3</sup> In die HO-Gruppe wurden nur Kinder mit deutlichen Befunden in mindestens zwei neurodiagnostischen Indikatorgruppen (Anamnese, neurologische Unters., EEG) aufgenommen. Die neurologische Untersuchung umfaßte die körperliche Abklärung des Neurostatus und Laboruntersuchungen wie Echoenzephalogramm, PEG, Liquor- und Röntgenuntersuchung, Szinti- und Arteriographie. Cerebralparetiker, Kinder mit chromosomalen Aberationen und metabolischen Störungen (z. B. Phenylketonurie PKU) wurden ausgeklammert.

stiziert wurden. Durch die Parallelisierung arbeiteten wir konservativ und unterschätzten die Unterschiede zwischen den Gruppen in den Hirnschadentests. Da wir sowohl die Facharzt diagnose als auch die Testkombinationen als zwei Indikatorgruppen *einer*<sup>4</sup> latenten Variablen („Hirnschaden“) ansehen, vermuten wir, kongenerische Tests vor uns zu haben. Aus diesem Grunde ist wegen der Fehlerbelastung *beider* Indikatorgruppen keine 100%ige Trefferquote zu erwarten. Auf jeden Fall sehen wir beide Informationsquellen apriori als gleichwertig an (Paine, 1965; Burgess et al., 1970; Zimet & Fishman, 1970).

Alle Probanden wurden in derselben Reihenfolge mit dem Hawik, Bender-Gestalt-Test (BGT), Background Interference Procedure (BIP)<sup>5</sup> untersucht. BGT und BIP wurden nach dem von Schlange et al. (1972) im Göttinger Formreproduktionstest (GFT) verwendeten System ausgewertet, so daß BGT und GFT identisch sind.

### 3. Datenanalyse und Ergebnisse

#### 3.1 univariater Ansatz

Die Ergebnisse der konventionellen Mittelwertvergleiche zeigen signifikante Unterschiede im BGT und in der BIP (Tabelle 1). Die Verschlechterungen der Ergebnisse

Tabelle 1  
Fehlermittelwerte im BGT und BIP für alle Gruppen  
t-Tests für abhängige Stichproben

	BGT	BIP
Normale	$\bar{x}_1 = 18.3$	$\bar{x}_2 = 19.3$
	$t = 2.21^*$ $df = 47$ $p = .0151$	
t-Tests für unabhängige Stichproben	$t = -3.57^{**}$ $df = 94$ $p = .0004$	$t = -6.14^{**}$ $df = 94$ $p = .0000$
Hirnorganiker	$\bar{x}_1 = 23.2$	$\bar{x}_2 = 27.9$
	$t = 10.50^{**}$ $df = 47$ $p = .0000$	
Gesamtstichprobe	$\bar{x}_1 = 20.7$	$\bar{x}_2 = 23.6$
	$t = 7.73^{**}$ $df = 47$ $p = .0000$	

\* = signifikant auf 5 % Niveau; \*\* = signifikant auf 1 % Niveau. Alle Irrtumswahrscheinlichkeiten p gelten für einsichtige Fragestellungen.

4 Wir differenzieren hier nicht zwischen verschiedenen Hirnschädigungen.

5 Bei der BIP werden im Gegensatz zum BGT oder GFT die nachzuzeichnenden Figuren nicht auf normalweißes, sondern auf ein mit Schlangenlinien bedrucktes Papier (s. a. Wallasch & Möbus, 1977) gezeichnet. Dieses Papier hat größere – allerdings theoretisch noch nicht erklärte – Effekte auf Hirnorganiker als auf Normale.

vom BGT zum BIP sind bei den Hirnorganikern erwartungsgemäß größer als bei der „normalen“ Kontrollgruppe. Alle Mittelwertsunterschiede liegen in den erwarteten Richtungen und befinden sich im Einklang mit der Literatur (Koppitz, 1964; Hall et al., 1968; Masini, 1969; Mecke, 1969; Adams, 1969, 1970; Yulis, 1970; Bauman & John, 1971; Kenny, 1971; Kropp & Cohen, 1972; Canter, 1968, 1970, 1976). Da signifikante Mittelwertsunterschiede noch keinen Aufschluß über die individualdiagnostische Güte des Verfahrens geben, haben wir die Gesamttrefferquoten und die bedingten Trefferwahrscheinlichkeiten für BGT und BIP berechnet (s. Tabelle 3, Prädiktionsregeln 1., 2.). Auch hier zeigt sich eine deutliche Überlegenheit der BIP mit einer punktbiserialen Validität von .54 und einer Gesamttrefferquote von 76 %. Zur Vereinheitlichung der cut-off Werte haben wir für jeden Prädiktor oder jede Prädiktorkombination eine Regression mit einer dichotomen Gruppenzugehörigkeitsvariablen (HO:  $y = 1$ ; N:  $y = 0$ ) als abhängiger Variablen gerechnet. Die Koeffizienten dieser Regression bezeichnen wir als ‚unnormiert‘. Sie können so normiert werden, daß die Mittelwerte der geschätzten Kriteriumswerte für die Gruppe der Hirnorganiker den Wert 1 und für die Gruppe der Normalen den Wert 0 annehmen. Obwohl zwei Parameter (a und b) geschätzt werden, haben sie für die Trefferquoten und Trefferwahrscheinlichkeiten keine verändernde Wirkung, da in diesem Fall Regression, Diskriminanzanalyse und lineare Skalentransformation identisch sind.

Ein erster Versuch, die beiden Testergebnisse nichtlinear zu kombinieren, erfolgt in Prädiktionsregel 3: Als HO gilt ein Kind eher, wenn es *sowohl* im BGT (Prätest) *als auch* in der BIP (Posttest) einen hohen Wert besitzt. Das Ergebnis dieses Ansatzes überzeugt uns aber nicht.

### 3.2 Auf Veränderungsmessungen basierender Ansatz

Als nächstes wurde der Vorschlag von Canter aufgegriffen, die konkurrente Validität des D-Wertes (= BIP – BGT)<sup>6</sup> zu überprüfen. Auch dieses Ergebnis überzeugte nicht (Prädiktionsregel 4). Als Hauptursache für die schlechte Validität des Differenzwertes vermuteten wir seine Abhängigkeit vom Ausgangsniveau (BGT). Sollte gar in der HO-Gruppe ein stärkerer Regressionseffekt<sup>7</sup> zu beobachten sein als in der N-Gruppe, würde der D-Wert für einen Teil der HO-Gruppe völlig invalide Werte liefern. Zur Über-

6 Vielfach wird in der Literatur vor der Verwendung von Differenzen bei Veränderungsmessungen gewarnt. Hauptgrund für diese Warnung ist die ‚mangelnde‘ Reliabilität der Differenzen. Schätzen wir z. B. die innere Konsistenz von BGT und BIP mit .95 (die Korrelation zwischen BGT und BIP beträgt .89), ergibt sich nach Lord (1963) für die Differenzen eine Reliabilität von .55. Diese niedrige Reliabilität schließt jedoch die Verwendung der D-Werte zur Prognose von *Gruppenzugehörigkeiten* nicht aus, denn: „... even though the mean change might be meaningful, reliability of the change scores would be meaningless if each subject made the same obtained gain (or loss) over the test – retest interval. In effect, then, it is differential, or relative, change among individuals ... that will be assessed by the reliability estimate. The estimate is not available when only a single total change score is considered (Webster & Bereiter, 1963, S. 48)“. Diese Überlegung trifft auch auf den für uns relevanten Fall zweier Gruppenchangescorers zu. Für uns gilt: Solange die mangelnde Reliabilität der Differenzen nur zu Vertauschungen der Pbn *innerhalb* aber nicht *zwischen* den Gruppen führt, ist unsere Zielsetzung – die Prognose der Gruppenzugehörigkeit – nicht ernstlich gefährdet.

7 bzw. ceilingeffekt, der nicht durch das Meßinstrument hervorgerufen wird.

prüfung dieser Hypothese haben wir den Verschlechterungsprozeß vom BGT zur BIP gruppenspezifisch mit einfachen linearen inhomogenen Differentialgleichungen (Tabelle 2) untersucht.

Beide Gruppen weisen signifikant schlechtere Testergebnisse in der BIP auf. Wir führen das auf die größere Schwierigkeit S (Schlangenpapier!) der BIP zurück. Der Abfall in der Testleistung ist dabei bei den HO erheblich stärker als bei den N.

Zur Untersuchung der gruppenspezifischen Verschlechterungsprozesse sehen wir BGT und BIP auf einer latenten Schwierigkeitsdimension S angeordnet an. Wir stellen die Hypothese auf, daß die Verschlechterungsrate  $dF/dS$  vom Fehlerniveau F abhängt. Bei 0 Fehlern im BGT soll dennoch eine Verschlechterung zu erwarten sein. Diese konstante Rate beträgt a. Die verbalen Hypothesen werden formal in eine einfache Differentialgleichung abgebildet:

$$(1) \quad \frac{dF}{dS} = b_j F + a_j$$

F = Fehlerzahl, S = Schwierigkeit der Tests (BGT und BIP stellen zwei Ausprägungen dieser Var. dar), j = Gruppenindex für HO und N.

Nur wenn sich herausstellt, daß  $b_N = b_{HO}$  und  $a_N \ll a_{HO}$  ist, hat es Sinn, die Differenz  $D = BIP - BGT$  allein als Prädiktionsmaß zu benutzen. Ist dagegen  $b_{HO} \ll b_N \leq 0$ , wird die Verschlechterung bei den Hirnorganikern stärker gedämpft als bei den Normalen. Die Differenzen fallen bei höheren Fehlerniveaus (und damit bei den Hirnorganikern) bei negativem b immer kleiner aus, da die Komponente bF die Komponente a bei hohem F dominiert.

Die Differentialgleichung wird standardmäßig gelöst (Smirnow, 1966, II, 8-17) und ergibt das allgemeine Integral:

$$(2) \quad F = e^{fbdS} [C + \int a e^{-fbdS} dS] = \frac{a}{b} (e^{bs} - 1) + F_0 e^{bs}$$

Als Randbedingung führen wir ein: für  $s = 0$  (BGT) ist  $F = F_0$  (Fehlerwert im BGT). Ordnet man dem BGT die Schwierigkeit  $s = 0$  und der BIP  $s = 1$  zu, ergibt sich:

$$(3) \quad F_1 = \frac{a}{b} (e^b - 1) + F_0 e^b = a^* + F_0 b^*$$

Coleman (1971) schlägt vor, die Parameter  $a^*$ ,  $b^*$  mit der Methode der kleinsten Quadrate zu schätzen und die fundamentalen Wandelparameter a und b zurückzurechnen nach:

$$(4) \quad b = \ln b^* \qquad a = a^* \ln b^* / (b^* - 1)$$

Die Schätzung nach der Methode der kleinsten Quadrate ist erlaubt, wenn die Autokorrelation der Fehler vernachlässigbar ist und die Prädiktoren meßfehlerfrei sind. Da die Konsistenz des BGT über der quadr. Korrel.  $r^2_{BGT/BIP} = .89^2$  liegen muß, nehmen wir letzteres als relativ gut approximiert an. Die Ergebnisse haben wir in Tabelle 2 dargestellt. Zur Prüfung auf gruppenspezifische Veränderungen wurde ein Test auf Gleichheit der Regressionsparameter (Test auf „Strukturbruch“) gerechnet (Schneeweiss, 1971). Danach weichen die Regressionsparameter der Gruppenregressionen signifikant von der Gesamtregression ab: Die Veränderungsprozesse sind nach diesem Modell ver-

Tabelle 2  
BIP/BGT-Regressionen und zugehörige Differentialgleichungen

Gruppe	Regression	$r_{\text{BIP/BGT}}$	Differentialgleichungen	$-\frac{a}{b}$
	$\text{BIP} = b \cdot \text{BGT} + a^*$		$\frac{dF}{dS} = b F + a$	
Normale	$\text{BIP} = .9436\text{BGT} + 2.0305$	0.89	$\frac{dF}{dS} = -.0581 F + 2.0900$	36.
Hirnorganiker	$\text{BIP} = .8930\text{BGT} + 7.2073$	0.89	$\frac{dF}{dS} = -.1131 F + 7.6228$	67.
Gesamtstichprobe	$\text{BIP} = 1.017\text{BGT} + 2.4928$	0.89	$\frac{dF}{dS} = +.0178 F + 2.4707$	
Test auf Strukturbruch: $F_{9,2}^2 = 18.83^{**}$ ( $p = .0000$ )			** signif. auf 1 % Niveau	

schieden geartet. Die konstante Veränderungsrate  $a$  ist bei den HO erwartungsgemäß größer als bei den N. Die Dämpfung ist bei den HO doppelt so groß wie bei den Normalen. Dieses muß sich negativ auf die Validität der D-Werte niederschlagen.

Nimmt man auch weiterhin identische Prozesse innerhalb jeder Gruppe, Konstanz der Parameter  $a$ ,  $b$ , Gültigkeit des linearen Ansatzes an, kann man die hypothetischen Gleichgewichtspunkte  $-a/b$  bei wachsender Schwierigkeit  $S$  des Tests ermitteln. Diese pendeln sich bei den N bei 36 Fehlern und bei den HO bei 67 Fehlern ein: zunehmender Übungsgewinn und zunehmende Testschwierigkeit lassen nach diesem hypothetischen Gedankenmodell die Leistung im Test gegen diese Gleichgewichtswerte konvergieren. In einer empirischen Überprüfung dieser Hypothese müßte man aber sicher die eine oder andere Annahme zu Gunsten einer weniger restringierenden verwerfen.

### 3.3 Auf nichtlinearer Kombination beruhende Ansätze

In 3.2 wurde angeregt, das Ausgangsniveau bei der Verwendung der D-Werte zu berücksichtigen. Wir haben die Prädiktionsregel 5 entsprechend konzipiert. Die punktbiseriale Validität liegt mit .58 dann auch höher als bei den anderen Regeln. Als HO gilt eher derjenige, der *sowohl* im BGT einen hohen Wert besitzt *als auch* eine große Verschlechterung vom BGT zum BIP aufweist. Die entsprechende Regel unter Verwendung der BIP (Regel 6) stellt dagegen keinen Informationsgewinn dar.

Die Prädiktionsregeln 7. und 8. sind statistisch konservativ aufgebaut. Nur wenn ein Pb im BGT *und* in der BIP einen hohen Fehlerwert *und* eine große Verschlechterung aufweist, wird er eher als HO bezeichnet (Regel 7). Die Trefferquote ist der von Regel 5 vergleichbar. Erfreulicherweise steigt die bedingte Trefferwahrscheinlichkeit  $P$  [Diagnose, Hypothese = HO | Test = HO] auf .83 an und fällt die bedingte Fehlerwahrscheinlichkeit  $P$  [Test = HO | Diagnose, Hypothese = N] auf .13. Quadriert man  $D$ , so daß die Varianz der  $D$  der des BGT oder der der BIP angenähert wird, erhöht sich die bedingte Trefferwahrscheinlichkeit sogar auf .89 (Regel 8) und fällt die bedingte Fehlerwahrscheinlichkeit auf .06. Leider wird die Wahrscheinlichkeit  $P$  [Diagnose = N | Test

Tabelle 3

## Ergebnisse der verschiedenen linearen und nichtlinearen Prädiktionsregeln

Prädiktionsregel (= Diskriminanzfunktion bzw. Regression) und Mittelwerte des Prädiktors	Trefferquoten und bedingte Trefferwahrscheinlichkeiten		Klassifikationstabelle (Hypothese vs. Test) beim cutoff $\hat{y}_i = 0.5$		unnormierte und normierte Koeffizienten b und a																		
	$r_{yy}$	P[Diag=N Test=N]	P[Diag=HO Test=HO]	Test		b	a																
1. $Y_i = b X_{1i} + a + e_i$ $X_1 = \text{BGT}$ $\bar{x}_N = 18.2708$ $\bar{x}_{HO} = 23.1667$	.35	64%:36%	= .62	= .66	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>34</td><td>14</td><td>48</td></tr> <tr><td>HO</td><td>21</td><td>27</td><td>48</td></tr> <tr><td></td><td>55</td><td>41</td><td>96</td></tr> </table>			N	HO	N	34	14	48	HO	21	27	48		55	41	96	0.24421135 (unnormiert)	-.005975384 (normiert)
	N	HO																					
N	34	14	48																				
HO	21	27	48																				
	55	41	96																				
2. $Y_i = b X_{2i} + a + e_i$ $X_2 = \text{BIP}$ $\bar{x}_N = 19.2708$ $\bar{x}_{HO} = 27.8958$	.54	76%:24%	= .78	= .75	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>35</td><td>13</td><td>48</td></tr> <tr><td>HO</td><td>10</td><td>38</td><td>48</td></tr> <tr><td></td><td>45</td><td>51</td><td>96</td></tr> </table>			N	HO	N	35	13	48	HO	10	38	48		45	51	96	.033197904 (unnormiert)	.282917246 (normiert)
	N	HO																					
N	35	13	48																				
HO	10	38	48																				
	45	51	96																				
3. $Y_i = b X_{1i} X_{2i} + a + e_i$ $\bar{x}_N = 391.479$ $\bar{x}_{HO} = 687.792$	.42	67%:33%	= .64	= .70	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>36</td><td>12</td><td>48</td></tr> <tr><td>HO</td><td>20</td><td>28</td><td>48</td></tr> <tr><td></td><td>56</td><td>40</td><td>96</td></tr> </table>			N	HO	N	36	12	48	HO	20	28	48		56	40	96	.000595437 (unnormiert)	.178681086 (normiert)
	N	HO																					
N	36	12	48																				
HO	20	28	48																				
	56	40	96																				
4. $Y_i = b D_i + a + e_i$ $\bar{x}_N = 1.00000$ $\bar{x}_{HO} = 4.72917$	.52	69%:31%	= .70	= .67	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>31</td><td>17</td><td>48</td></tr> <tr><td>HO</td><td>13</td><td>35</td><td>48</td></tr> <tr><td></td><td>44</td><td>52</td><td>96</td></tr> </table>			N	HO	N	31	17	48	HO	13	35	48		44	52	96	.071529067 (unnormiert)	.295099027 (normiert)
	N	HO																					
N	31	17	48																				
HO	13	35	48																				
	44	52	96																				
5. $Y_i = b X_{1i} D_i + a + e_i$ $\bar{x}_N = 15.91667$ $\bar{x}_{HO} = 104.5833$	.58	74%:26%	= .72	= .77	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>38</td><td>10</td><td>48</td></tr> <tr><td>HO</td><td>15</td><td>33</td><td>48</td></tr> <tr><td></td><td>53</td><td>43</td><td>96</td></tr> </table>			N	HO	N	38	10	48	HO	15	33	48		53	43	96	.003800149 (unnormiert)	.271041005 (normiert)
	N	HO																					
N	38	10	48																				
HO	15	33	48																				
	53	43	96																				
6. $Y_i = b D_i X_{2i} + a + e_i$ $\bar{x}_N = 26.50000$ $\bar{x}_{HO} = 136.479$	.55	71%:29%	= .67	= .75	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>38</td><td>10</td><td>48</td></tr> <tr><td>HO</td><td>18</td><td>30</td><td>48</td></tr> <tr><td></td><td>56</td><td>40</td><td>96</td></tr> </table>			N	HO	N	38	10	48	HO	18	30	48		56	40	96	.002745264 (unnormiert)	.276289575 (normiert)
	N	HO																					
N	38	10	48																				
HO	18	30	48																				
	56	40	96																				
7. $Y_i = b X_{1i} D_i X_{2i} + a + e_i$ $\bar{x}_N = 480.5417$ $\bar{x}_{HO} = 3165.04$	.56	74%:26%	= .69	= .83	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>42</td><td>6</td><td>48</td></tr> <tr><td>HO</td><td>19</td><td>29</td><td>48</td></tr> <tr><td></td><td>61</td><td>35</td><td>96</td></tr> </table>			N	HO	N	42	6	48	HO	19	29	48		61	35	96	.000114964 (unnormiert)	.290445040 (normiert)
	N	HO																					
N	42	6	48																				
HO	19	29	48																				
	61	35	96																				
8. $Y_i = b X_{1i} D_i^2 X_{2i} + a + e_i$ $\bar{x}_N = 3549.167$ $\bar{x}_{HO} = 21219.7$ $D_i = \text{BIP}_i - \text{BGT}_i$	.42	74%:26%	= .67	= .89	<table border="1"> <tr><td></td><td>N</td><td>HO</td></tr> <tr><td>N</td><td>45</td><td>3</td><td>48</td></tr> <tr><td>HO</td><td>22</td><td>26</td><td>48</td></tr> <tr><td></td><td>67</td><td>29</td><td>96</td></tr> </table>			N	HO	N	45	3	48	HO	22	26	48		67	29	96	.000009612 (unnormiert)	.380957809 (normiert)
	N	HO																					
N	45	3	48																				
HO	22	26	48																				
	67	29	96																				

Die Berechnung der unnormierten Koeffizienten  $b_0, a_0$  erfolgte über eine Regression mit einer dichotomisierten abhängigen Variablen  $y$  ( $HO: y = 1, N: y = 0$ ) und mit einer unabhängigen Variablen  $x = f(\text{BGT}, \text{BIP})$ . Die normierten Koeffizienten  $b_1, a_1$  errechnen sich einfach aus:  $b_1 = 1/(\bar{x}_{HO} - \bar{x}_N)$  und  $a_1 = -\bar{x}_N/(\bar{x}_{HO} - \bar{x}_N)$ . Mit dieser Normierung ist gewährleistet, daß  $\hat{y}_{HO} = 1.0$  und  $\hat{y}_N = 0.0$ . Der cutoff 0.5 bleibt davon unberührt.



= N] und damit die Validität ( $r = .42$ ) gedrückt. Dieses ist durch die Nichtsymmetrie der Verteilung in der vierfeldrigen Klassifikationstabelle bedingt. Jensen (1973) wies in einer Diskussion über funktionale Hierarchien auf die Wichtigkeit solcher Verteilungen hin. Die Hirnschädigung ist eine notwendige aber nicht hinreichende Voraussetzung, um im Test als HO aufzufallen. Normale werden im Test nicht „aus Versehen“ als HO qualifiziert.

Fällt jemand im Test als HO auf, kann man einigermaßen sicher sein, daß die Diagnose entsprechend ist.

Es stellt sich die Frage, ob man nicht durch eine einfache Verschiebung des cutoffs bei der BIP ähnlich hohe Trefferwahrscheinlichkeiten  $P[\text{Diag} = \text{HO} \mid \text{Test} = \text{HO}]$  wie bei Regel 7 oder 8 erzielen kann. Wie aus Tabelle 4 ersichtlich wird dabei die Gesamt-

Tabelle 4

Wahl verschiedener cutoffs und die damit für die Trefferquote verbundenen Effekte

Prädiktionsregel (= Diskriminanzfunktion bzw. Regression oder Skalentransformation)	cutoff auf unnommierter bzw. normierter Funktion		Gesamttrefferquote	bedingte Trefferwahrscheinlichkeiten	
	unnormiert	normiert		$P[\text{Diag}=\text{N} \mid \text{Test}=\text{N}]$	$P[\text{Diag}=\text{HO} \mid \text{Test}=\text{HO}]$
2. $Y_1 = b X_{2i} + a + e_i$	0.50	0.500	76%:24%	.78	.75
	0.55	0.675	70%:30%	.66	.76
	0.60	0.850	69%:31%	.65	.76
	0.65	1.024	67%:33%	.61	.81
	0.70	1.198	62%:38%	.58	.80
	0.75	1.371	60%:40%	.56	.81
8. $Y_1 = bX_{1i}D_1X_{2i} + a + e_i$	0.50	0.50	73%:27%	.67	.89
	0.475	0.35	72%:28%	.67	.82
	0.45	0.205	68%:32%	.64	.73
	0.40	-0.089	68%:32%	.74	.63

trefferquote drastisch reduziert, so daß Regel 7 oder 8 diesem Vorgehen gegenüber überlegen sind.

Zur Veranschaulichung der Regeln 7 und 8 sind in Figur 1 die  $P_{bn}$  und die Linien gleicher Zugehörigkeitswahrscheinlichkeit (= optimale Grenzlinie) im Testraum, der durch BGT und BIP aufgespannt wird, eingezeichnet. Die Grenzen wurden folgendermaßen bestimmt. Wir setzten den cutoff 0.5 in die Gleichung der Regel ein

$$(5) \quad 0.0 = f(\text{BGT}, \text{BIP}) - 0.5$$

und bestimmten die Nullstellen<sup>8</sup> des Polynoms für einen bestimmten festgehaltenen BGT-Wert.

Zusammenfassend kann man sagen, daß unter Anlegung des konventionellen Gütekriteriums „Gesamttrefferquote“ Regeln 2, 5, 7 und 8 zur Auswahl stehen, wobei wegen der einfacheren und schnelleren Anwendung Regel 2 der Vorzug einzuräumen ist. Berücksichtigt man dagegen die von uns für wichtig gehaltene Maximierung von  $P[\text{Diag}$

<sup>8</sup> Unter den mehreren möglichen Nullstellen natürlich nur die sinnvollen.

**FIGUR 1:** Bivariater Testraum (BGT x BIP) und optimale Grenzen für Prädiktionsregeln 7 und 8

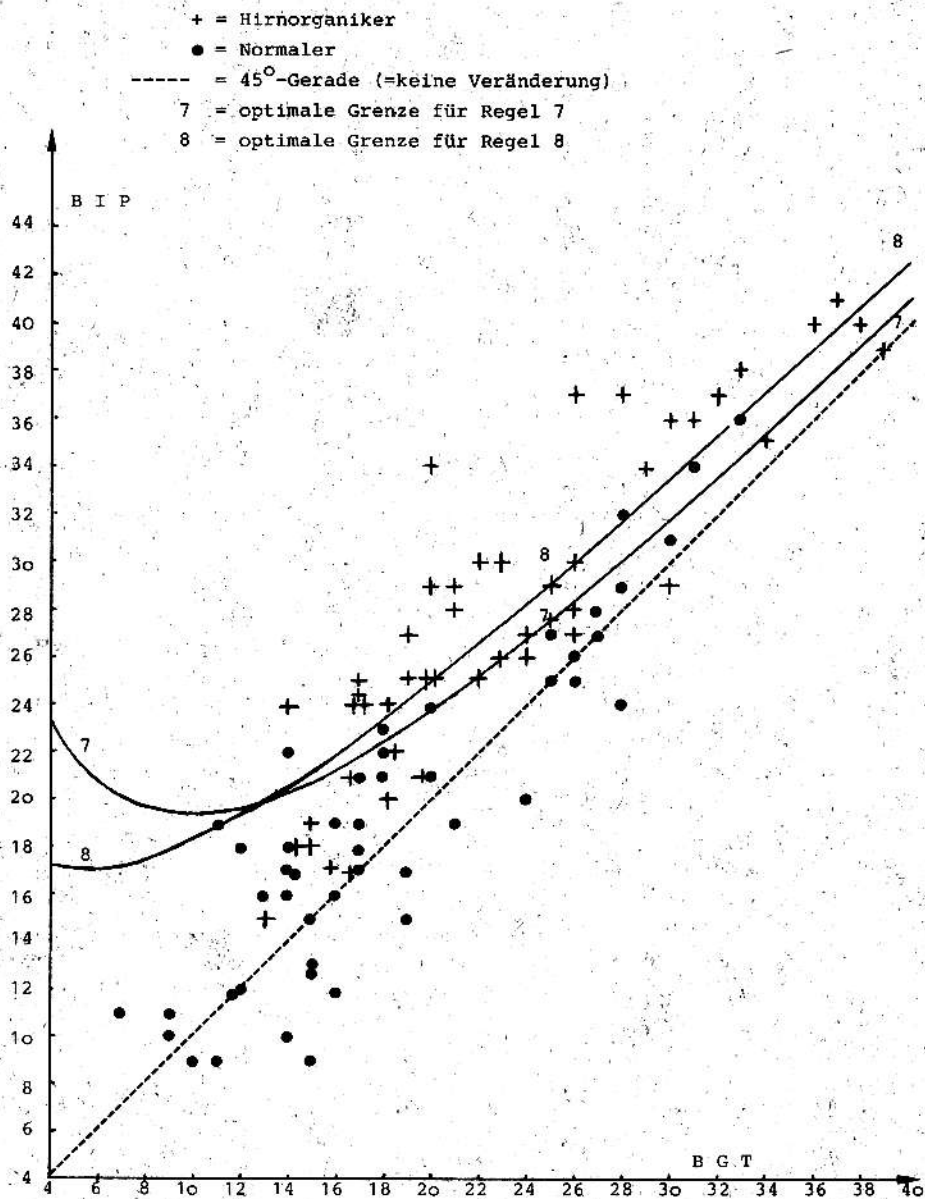


Abb. 1

Bivariater Testraum (BGT x BIP) und optimale Grenzen für Prädiktionsregeln 7 und 8

= HO | Test = HO] und die Minimierung von  $P[\text{Test} = \text{HO} | \text{Diag} = \text{N}]$ , so scheinen die nichtlinearen Regeln 7 und 8 überlegen zu sein.<sup>9</sup>

Die in Tabelle 3 aufgeführten Trefferquoten lassen jedoch keine verlässliche Beurteilung der Güte einer Diagnoseregeln zu, weil sie zu optimistisch sind, eine Kreuzvalidierung fehlt und die Stabilität der Entscheidungen nicht untersucht wurde. So können z. B. zwei Diagnoseregeln die gleichen Trefferquoten aufweisen – trotz unterschiedlicher Schätzfehlervarianzen. Wir wollen diese Fragen im folgenden Abschnitt mit der Jackknife-Methode von Tukey (1958) – einem Spezialfall der random subsample replication (RSSR) Techniken (Finifter, 1972) – einer Klärung näher bringen.

#### 4. Kreuzvalidierung der Trefferquoten und Jackknife-Schätzung der Parameter

Bei multivariaten Verfahren ist die Notwendigkeit der Kreuzvalidierung der Ergebnisse weithin akzeptiert, weil diese komplexen statistischen Verfahren Fehler bzw. Daten- und Stichprobenindiosynkrasien bei der Parameterbestimmung einseitig ausschöpfen (Cureton, 1951; Frank, Massy & Morrison, 1965; Herzberg, 1969; Morrison, 1969; Nicholson, 1960; Thorndike, Weiss & Davis, 1968; Wallasch & Möbus, 1976; Wherry, 1951). Um diesem Effekt zu entgehen, haben wir unsere Variablen- und Parameterzahl drastisch reduziert. Zur realistischen Beurteilung der Trefferquoten haben wir folgenden Weg gewählt. Die virtuelle Replikation der Studie zur Erhebung „frischer“ Daten mußte leider wegen des zeitlichen und organisatorischen Aufwandes ausgeschlossen werden. Ebenso konnte der Weg der Stichprobenhalbierung zur Parameterschätzung und Kreuzvalidierung wegen des kleinen Umfangs ( $N = 96$ ) nicht besritten werden. Die virtuelle Replikation wurde deshalb zugunsten von Pseudoreplikationen (Mosier, 1951) zurückgestellt. Dieses geschah mit der „single-cross-validation“ bzw. „leave-out-one“-Technik von Mosteller & Tukey (1968) und Lachenbruch (1965, 1975). In ihr werden ebenso wie bei der „jackknife“-Methode (Quenouille, 1956; Tukey, 1968; Brillinger, 1964; Miller, 1964; Dempster, 1966; Mosteller & Tukey, 1968; Frankel, 1971) aus dem Datensatz sukzessiv nach einem bestimmten Schema kleine Stichproben entfernt und mit dem (weitaus größeren) Rest Parameter geschätzt. Der kleinere Rest dient der Kreuzvalidierung. Da ständig nach einem bestimmten Schema Suchstichproben gezogen werden, kann man die verschiedenen Methoden als Spezialfälle der random subsample replication (RSSR) Methoden ansehen, die ein empirisches Gegenstück zur statistischen Testtheorie darstellen (Finifter, 1972).

##### 4.1 Kreuzvalidierung und Schätzung der Trefferquoten

Man kann je nach Annahmen und Voraussetzungen mindestens fünf verschiedene Fehlerraten berechnen (Lachenbruch, 1975; Hills, 1966; Fukunaga

---

<sup>9</sup> Wir sind in unserer Untersuchung von gleichen apriori-Wahrscheinlichkeiten ausgegangen. Nach unseren Beobachtungen wächst die apriori-Wahrscheinlichkeit für die HO, je mühseliger die Testauswertung für den Praktiker wird, weil er nur noch testet, wenn er schon z. B. 70:30 „sicher“ ist, einen HO vor sich zu haben.

& Kessel, 1971: 1. optimum, 2. actual, 3. expected actual, 4. estimated actual, 5. apparent error rate. In Tabelle 3 sind die apparent error rates, die sich auf den Prozentsatz der richtig klassifizierten Stichprobenmitglieder beziehen, angegeben. Die Klassifikation wird mit den an derselben Stichprobe ermittelten Parametern durchgeführt. Diese Fehlerratenbestimmung zeichnet sich gegenüber den anderen vier Verfahren durch einfache Berechnung und Freiheit von Verteilungsannahmen aus. Als Nachteil, der sich besonders schwerwiegend in kleinen Stichproben auswirkt, muß ihre mangelnde Erwartungstreue angeführt werden: Die Schätzungen fallen generell zu hoch aus. Da wir auf Verteilungsannahmen verzichten wollen (keine Homogenität der Gruppenvarianzen bei den Prädiktoren, kleine Stichprobe) schätzen wir die expected actual error rate mit der leave-out-one Methode (Lachenbruch, 1975, S. 29 ff.), die ebenfalls den Vorteil der Verteilungsfreiheit besitzt.

Die Daten der Person  $i$  werden aus dem Gesamtdatensatz  $D_g$  eliminiert. Mit dem um  $i$  verminderten Datensatz  $D_{g(i)}$  wird eine Diskriminanzanalyse (hier: Regression) gerechnet. In diese Gleichung werden die Daten  $x_i$  der Person  $i$  eingesetzt. Die Prognosen  $\hat{y}_i = f(x_i, D_{g(i)})$  mit  $i = 1, 2, \dots, 96$  werden dann in Klassifikationstabellen umgesetzt und anschließend die Trefferquoten und bedingten Trefferwahrscheinlichkeiten berechnet (s. Tabelle 5). Erfreulicherweise fallen die Trefferquoten nur bis zu 1 % und die bedingten Trefferwahrscheinlichkeiten bis zu 0.03 ab. Hier zahlt sich die Beschränkung in der Parameterzahl aus.

Tabelle 5  
Ergebnisse der leave-out-one Technik

Prädiktionsregel	Gruppenmittelwerte		Koeffizienten		gejacknifete Koeffizienten		Standardmeßfehler		t-Werte		Signif. niv.	
	$\bar{x}_N$	$\bar{x}_{HO}$	a	b	a*	b*	$s_{a^*}$	$s_{b^*}$	$t_{a^*}$	$t_{b^*}$	$P_{t_{a^*}}$	$P_{t_{b^*}}$
1. BGT	18.3	23.2	-3.731915	.2042553	-3.397847	.1879938	1.194	.058	-2.8	3.2	.00559	.00197
2. BIP	19.5	27.9	-2.314356	.1188119	-2.234544	.1154471	.481	.019	-4.6	5.7	.00007	.00001
3. BGT*BIP	391.5	687.8	-1.321170	.0033748	-1.236624	.0032039	.390	.001	-3.1	4.2	.00243	.00019
4. D	1.0	4.7	-2.681564	.2681564	-2.453464	.2601827	.158	.046	-1.6	5.6	.12000	.00001
5. BGT*D	15.9	104.6	-0.179511	.0112782	-0.167675	.0110377	.108	.002	-1.6	6.7	.11900	.00000
6. D*BIP	26.5	136.5	-0.240954	.0090926	-0.228600	.0088631	.104	.001	-2.2	6.1	.02808	.00000
7. BGT*D*BIP	480.5	3165.0	-0.179006	.0003725	-0.170126	.0003634	.083	.000	-2.1	6.2	.04055	.00000
8. BGT*D <sup>2</sup> *BIP	3549.2	21219.7	-0.200853	.0000566	-0.187953	.0000533	.067	.000	-2.8	3.8	.00604	.00055

Prädiktionsregel	Klassifikationstabelle		Trefferquoten mit $\hat{y}_i = f(x_i, D_{g(i)})$	P[Diagnose = N   Test = N]	P[Diagnose = HO   Test = HO]
	N	HO (Test)			
1. BGT	N	34	63 % : 37 %	.61	.65
	HO	22			
2. BIP	N	34	75 % : 25 %	.77	.73
	HO	10			
3. BGT*BIP	N	36	68 % : 32 %	.65	.71
	HO	19			
4. D	N	35	69 % : 31 %	.67	.70
	HO	17			
5. BGT*D	N	38	74 % : 26 %	.72	.77
	HO	15			
6. D*BIP	N	38	70 % : 30 %	.67	.74
	HO	19			
7. BGT*D*BIP	N	41	73 % : 27 %	.68	.81
	HO	19			
8. BGT*D <sup>2</sup> *BIP	N	44	73 % : 27 %	.67	.87
	HO	22			

## 4.2 Jackknife-Schätzung der Parameter

Die Jackknife-Methode ist eine „dependent replicated subsampling“ Version der RSSR-Techniken, die meist nach der Datenerhebung eingesetzt wird. Sie geht auf Quenouille (1956) zurück, der den Bias bei der Schätzung von Zeitreihenkorrelationen verkleinern wollte. Tukey (1958) entwickelte die Methode weiter und taufte sie „Jackknife“-Methode, „to suggest the broad usefulness of a technique as a substitute for specialized tools that may not be available, just as the Boy Scout's trusty tool serves so variedly. The jackknife offers ways to set sensible confidence limits in complex situations. The basic idea is to assess the effect of each of the groups into which the data have been divided not by the result for that group alone, ..., but rather through the effect upon the body of data that results from omitting that group (Mosteller & Tukey, 1968, S. 134)“.

Drei Gedanken bilden die Grundlage des Jackknifes:

1. Ein Datensatz  $D_g$  kann nicht nur in zwei sondern in maximal  $k = N$  Teile aufgespalten werden.
2.  $k$  Substichproben werden zur Berechnung von transformierten Schätzungen herangezogen, die wie annähernd unabhängige, identisch verteilte, Beobachtungen behandelt werden.
3. Student's  $t$ -Verteilung kann auf diese Schätzungen angewendet werden zur Konstruktion von Konfidenzintervallen und zur Testung von Hypothesen (Brillinger, 1964; Miller, 1964).

Wir haben in unserer Untersuchung  $k = N$  gesetzt und das Stichprobenschema der leave-out-one Technik benutzt: jede eliminierte Stichprobe hat die Größe  $n = 1$  (wobei  $N = nk$ ). Für jeden Datensatz  $D_{g(i)}$  ( $i = 1, 2, \dots, 96$ ) wurden die Parameter  $a_{(i)}$ ,  $b_{(i)}$  geschätzt (alles in REAL\*8 auf IBM 370/168). Diese 96 Schätzungen sind zwar wegen der großen Datenüberlappung bei 94 Pbn hoch korreliert, wird der Datensatz doch jeweils nur um eine Person reduziert. Aber in den meisten Fällen können die mit der Grundformel (6) linear transformierten Schätzungen als unabhängig verteilte Pseudowerte  $a_{*i}$ ,  $b_{*i}$  angesehen werden. Die Pseudowerte spiegeln den Effekt, der sich aus der Entfernung einer Substichprobe aus  $D_g$  ergibt, wieder. Sie reflektieren die Störanfälligkeit von Parameterschätzungen gegenüber Zufallseinflüssen (random shocks) auf den zugrunde liegenden Datenkörper.

$$(6) \quad \begin{aligned} a_{*i} &= Na - (N-1) a_{(i)} \\ b_{*i} &= Nb - (N-1) b_{(i)} \end{aligned}$$

$a, b$  = Parameter der Gesamtregression;  $a_{(i)}, b_{(i)}$  = Regressionsparameter auf der Basis von  $D_{g(i)}$ ;  $a_{*i}, b_{*i}$  = Pseudowerte.

Mittelt man die individuellen Pseudowerte, erhält man die Jackknife-Schätzung, die eine annähernd erwartungstreue Schätzung des Populationsparameter ist:

$$(7) \quad \begin{aligned} a_* &= \sum_i a_{*i} / N \\ b_* &= \sum_i b_{*i} / N \end{aligned} \quad a_*, b_* = \text{Jackknife-Schätzungen}$$

Die Varianz  $S^2$  der Pseudowerte kann in den Standardschätzfehler überführt werden. Mit dessen Hilfe und der der t-Verteilung lassen sich dann Konfidenzintervalle und Signifikanztests durchführen: "The key idea is that, in a wide variety of problems, the pseudo-values can be used to set approximate confidence limits, using Student's t, as if they were the results of applying some complex calculation to each of k independent pieces of data. The words 'as if' are vital here; Student's t performs well in many circumstances where the pseudovalues deviate substantially from independence (Mosteller & Tukey, 1968, S. 135)".

Wir haben alle Berechnungen in Tabelle 5 zusammengefaßt. Die Koeffizienten sind alle bis auf die Konstanten  $a_*$  in Regel 4 - 7 auf dem 1 % Niveau signifikant von Null verschieden.

### 4.3 Stabilitätsschätzungen der Prognosen

In 4.1 wurden die Prädiktionsregeln f (a, b) kreuzvalidiert und in 4.2 der Bias herausgezogen. Will man beides kombinieren, d. h. die Jackknife-Schätzung kreuzvalidieren, und ermitteln, ob Prädiktionsregeln mit annähernd gleichen Trefferquoten und bedingten Trefferwahrscheinlichkeiten sich in der Sicherheit ihrer Aussagen unterscheiden, muß man die leave-out-one Technik zur leave-out-two Methode erweitern. Wir entfernen jetzt nicht nur die Person i sondern auch die Person j. Für jedes i erhalten wir 95 Pseudoregressionen:

$$(8) \quad \begin{aligned} a_{*j}(i) &= (N - 1) a_{(i)} - (N - 2) a_{(i, j)} \\ b_{*j}(i) &= (N - 1) a_{(i)} - (N - 2) a_{(i, j)} \end{aligned}$$

Pseudokoeffizienten für Pb i (j = 1, ..., 96; j ≠ i)

Setzen wir die Daten  $x_j$  des Pb i in die 95 Gleichungen mit den Koeffizienten (8) ein, erhalten wir 95 Prognosen  $\hat{y}_{*i(j)} = f(x_j, a_{*j}(i), b_{*j}(i))$  mit Mittelwert  $\bar{y}_{*i}$  und Standardschätzfehler  $s_{*i} = (\text{Var}(\hat{y}_{*i})/94)^{1/2}$ . Uns interessiert jetzt die Frage, wie weit die Prognose  $\bar{y}_{*i}$  vom cut-off 0.5 entfernt ist. Diese Distanz wird in Einheiten des Standardschätzfehlers gemessen und sollte im Sinne der Stabilität der Aussage möglichst groß sein. Hierzu berechnen wir die kritische Distanz  $z_{*i}$  (s. Tabelle 6). Die entsprechenden Kreuztabellen, Trefferquoten und bedingten Wahrscheinlichkeiten finden sich in Tabelle 7. Die Gesamttrefferquoten fallen maximal 2 % und die bedingten Wahrscheinlichkeiten maximal 0.03 ab. Anders sieht natürlich das Bild aus, wenn man die kritische Distanz zum cutoff in Rechnung zieht. Begnügt man sich mit der Aufteilung der Distanzen in vier Bereiche HO, HO?, N?, N, scheinen von der „sicheren“ Trefferquote (Aussagen werden nur getroffen, wenn  $|z_{*j}| > 3.0$ ) die Regel 5 (BGT\*D) und von der bedingten Trefferwahrscheinlichkeit P [Diag = HO | Test = HO] her die Regeln 4, 6, 7 überlegen zu sein. Eindeutiger wird das Bild, wenn man den „sicher“ richtigen Bereich (HO und N) ausdifferenziert. Wir haben für jede Regel die Produktsumme aus  $z_{*j}$  und Häufigkeit von  $z_{*j}$  berechnet. Danach ergibt sich folgendes Bild (s. Tabelle 8).

Tabelle 6

Ergebnisse der leave-out-two Technik  
 mit Standardmeßfehler  $s_{*j}$ ; standardisierte Distanzen  $z_{*1} = |\bar{y}_{*1} - .5| / s_{*1}$   
 zum cutoff und deren Häufigkeiten

		$\bar{y}_{*1}$ liegt auf der falschen Seite des cutoffs 0.5										$\bar{y}_{*1}$ liegt auf der richtigen Seite des cutoffs 0.5																	
		$z_{*1}$					$z_{*1}$					$z_{*1}$					$z_{*1}$												
		14-13	13-12	12-11	11-10	10-9	9-8	8-7	7-6	6-5	5-4	4-3	3-2	2-1	1-0	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14
BGT																													
Normale																													
HO		8	5	1	2	1	14	16																					
		6	11	5	1	2	19	4																					
BIP																													
N		4	2	2	2	2	1	2	4	13	14																		
HO		3	3	3	1	5	5	2	4	9	13																		
BGT*DIP																													
N		1	5	5	1	2	2	4	8	14																			
HO		9	7	1	2	1	2	4	8	14																			
D																													
N		2	2	6	7	4	4	6	10	11																			
HO		1	6	6	6	8	8	14	9																				
BGT*D																													
N		3	2	2	4	1	3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
HO		3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
D*BIP																													
N		3	4	2	1	2	3	2	2	1	1	5	2	1	1	4	2	4	26										
HO		3	4	2	4	4	1	5	2	4	3	17	3																
BGT*D*BIP																													
N		3	1	2	1	1	2	1	1	2	1	2	3	5	27														
HO		3	3	3	1	2	3	2	1	1	8	4	13	3															
BGT*D*BIP																													
N		5	2	7	2	1	2	2	1	2	3	5	14	3	3	1	1	21	7	5	4								
HO		5	2	7	2	1	2	2	1	2	3	5	14	3	3	1	1	21	7	5	4								

$\hat{y}_{*1(0)} = f(x_{*1}, D_{*1}(i))$

„sicher“ falscher Bereich

unsicherer Bereich

„sicher“ richtiger Bereich

Tabelle 7

Ergebnisse der leave-out-two Technik:  
 Unter Berücksichtigung von z<sub>ij</sub> Kreuzvalidierte Trefferquoten  
 und Trefferwahrscheinlichkeiten

Prädikationsregel	„sichere“ (a+h)% : (b+c+d+e+f)%		normale Trefferquote (a+b+g+h)% : (c+d+e+f)%		Klassifikationstabelle Test								P[Diag = N   Test = N] = P[ID = N   T = N]	P[ID = N   T = N + N']	P[ID = HO   T = HO]	P[ID = HO   T = HO + HO']	
	N	N?	H?	H	N	N?	H?	H	N	N?	H?	H					
1. BGT	21% : 79%	61% : 39%	N 16 HO 6	17 7 16 22	8 48 4 48								.73		.33	.60	.63
2. BIP	59% : 41%	77% : 23%	N 31 HO 9	3 6 1 12	8 48 26 48								.78		.76	.77	.73
3. BGT*BIP	51% : 49%	66% : 34%	N 34 HO 16	2 6 14 14	6 48 14 48								.68		.70	.64	.70
4. D	52% : 48%	69% : 31%	N 27 HO 7	4 13 6 12	4 48 23 48								.79		.85	.70	.67
5. BGT*D	63% : 37%	74% : 26%	N 33 HO 9	5 4 6 6	6 48 27 48								.79		.82	.72	.77
6. D*BIP	61% : 39%	70% : 30%	N 36 HO 13	2 7 6 6	3 48 23 48								.73		.88	.67	.74
7. BGT*D*BIP	60% : 40%	74% : 26%	N 38 HO 13	4 3 6 9	3 48 20 48								.75		.87	.69	.83
8. BGT*D*BIP	47% : 53%	73% : 27%	N 42 HO 21	3 3 2 2	0 48 3 48								.67	zu kleine Zellenbesetzung		.56	.89
			N a HO c	b f	d g								att+b*td ett+g*h				



Tabelle 8  
 Qualität der Prädiktionsregeln entsprechend der Sicherheit ihrer Aussage  
 (= kritische Distanz zum cutoff)

Prädiktionsregel	Produktsumme $\sum z_{*i} \text{ freq}(z_{*i})$	Rangplatz
1. BGT	60	8
2. BIP	247	5
3. BGT x BIP	172	7
4. D	214	6
5. BGT x D	304	3
6. D x BIP	288	4
7. BGT x D x BIP	324	2
8. BGT x D <sup>2</sup> x BIP	1034	1

Nach diesen Ergebnissen zeigt sich, daß eine erheblich größere Sicherheit in der Prognose erzielt werden kann, wenn man die Veränderungsmessungen in die Prognose integriert. Dieser integrative Ansatz ist den Einzelprädiktoren überlegen. Darüber hinaus ist es einleuchtend, daß die Sicherheit der Aussage steigt, wenn man die Hypothese  $H_0$  nur anstellt, wenn der  $P_b$  in mehreren Indikatoren gleichzeitig auffällt (Konjunktion von Indikatoren). Will man eine Empfehlung geben, würden wir die Regeln 8, 7, 5 für die Praxis empfehlen. Die Entscheidung darüber hängt aber auch von den Zielen des Untersuchers, seinen subjektiven Vermutungen über die Grundquoten etc. ab.

Die endgültige Beurteilung dieses vielversprechend erscheinenden Ansatzes, wie er in Regeln 8, 7, 5 vertreten wird, hängt natürlich von weiteren Untersuchungen mit virtuellen Replikationen und verschiedenartigen Stichproben ab. Besonders muß die Abgrenzung zu anderen klinischen Gruppen untersucht werden.

### 5. Prognosen in der klinischen Praxis

Man kann das in 4.3 beschriebene Verfahren verwenden, um in der Praxis die Stabilität und Sicherheit der Prognose abzuschätzen. Hierzu modifizieren wir nur (8) zu (9):

$$(9) \quad \begin{aligned} a_*(i) &= Na - (N - 1) a_{(j)} \\ b_*(i) &= Nb - (N - 1) b_{(j)} \end{aligned}$$

Die Prognose über die neu zu untersuchende  $P_b(i)$  wird in der klinischen Praxis folgendermaßen erstellt:

1. Aus dem Referenzdatensatz  $D_g$  wird jeweils die Person  $j$  entfernt ( $j = 1, 2, \dots, N$ ) und mit den  $(N - 1)$  verbleibenden Personen eine Regression oder Diskrimanzanalyse gerechnet. Wir erhalten Koeffizienten  $a_{(j)}, b_{(j)}$ .

2. Für jede eliminierte Person  $j$  berechne ich die Pseudokoeffizienten nach (9) und erhalte insgesamt  $N$  Pseudoregressionen oder Pseudodiskrimanzanalysen. Die Koeffizienten  $a, b$  berechne ich mit einer einzigen Regression oder Diskrimanzanalyse auf der Basis des Gesamtdatensatzes.

3. Mit den Daten  $x_i$  der Person  $i$  und den  $N$  Pseudoregressionen bzw. Pseudodiskriminanzanalysen (Koeffizienten aus (9) erstelle ich  $N$  Prognosen  $\hat{y}_{*i(j)}$ . Die Verteilung dieser Prognosen besitzt Mittelwert  $\bar{y}_{*i}$  und Standardschätzfehler  $s_{*i}$  (s. 4.3).

4. Der Mittelwert  $\bar{y}_{*i}$  besitzt die kritische Distanz  $z_{*i}$  zum cutoff. Ist diese kritische Distanz größer als 3.0, stellen wir eine Prognose. Bleibt sie darunter, suchen wir weitere klinische und aus der Verhaltensdiagnostik gewonnene Informationen über den  $P_b$   $i$  in die Prognose zu integrieren.

### Zusammenfassung

Wir versuchen bei der Prognose von hirnrorganischen Schäden einen Mittelweg zwischen multivariater und univariater Methodik zu beschreiten. Der multivariate Ansatz zeichnet sich zwar durch hohe Trefferquoten aus, birgt aber den Nachteil der schlechten Kreuzvalidierung in sich. Der univariate Ansatz läßt sich dagegen gut kreuzvalidieren, weist sich aber nicht durch sehr hohe Trefferwahrscheinlichkeiten aus. Die für eine Individualdiagnostik zu fordernde Präzision und Sicherheit soll mit einem nichtlinearen Ansatz erzielt werden, der Veränderungsmessungen einschließt. Eine Hirnschadensdiagnose wird nur dann gestellt, wenn mehrere Indikatoren simultan ansprechen (Konjunktion von Indikatoren). Wir verwenden bis zu drei Indikatoren (hohe Fehlerzahl im Prätest,<sup>1</sup> Verschlechterung von Prä- zum Posttest<sup>2</sup> und hohe Fehlerzahl im Posttest). Diese Indikatoren werden multiplikativ zu einer einzigen Variablen zusammengefaßt, die dann als Prädiktor in die Regression oder Diskriminanzanalyse eingeht.

Das Ergebnis dieses Ansatzes ist erfreulich. Die bedingten Trefferwahrscheinlichkeiten  $P$  [Diagnose = Hirnrorganiker | Test = Hirnrorganiker] liegen z. T. über .80 und lassen sich auch mit der leave-out-one Technik von Mosteller & Tukey und Lachenbruch kreuzvalidieren.

Während in der klassischen Diagnostik nur die Trefferquoten maximiert werden, versuchen wir in dieser Arbeit zusätzlich die *Sicherheit* der Prognose zu vergrößern. Die Sicherheit einer Aussage wächst mit der kritischen Distanz der Prognose zum cutoff. Es zeigt sich, daß Prognoseregeln mit ähnlichen Trefferquoten ganz erheblich in der Sicherheit ihrer Aussagen differieren.

### Summary

In the prediction of minimal brain damage we try to combine the advantages of the multivariate and univariate approaches. Though we observe high hit rates using a battery of tests, we have to realize the instability of the results when we start a crossvalidation. The univariate approach offers stability but shows a lack in precision which is absolutely undesirable in making prognoses for individuals. We try to avoid these disadvantages with a nonlinear approach containing measures of change. A prognosis is made only in the case of the simultaneous appearance of three indicators (conjunction of indicators). These indicators (high error rate in the pretest,<sup>1</sup> remarkable change from pre- to posttest,<sup>2</sup> high error rate in the posttest) are combined in a multiplicative way to a single predictor which can be put into a regression or discriminant equation.

1 Göttinger Formreproduktionstest von Schlange.

2 Background Interference Procedure von Canter.

The results are very promising. We observe a conditional hit probability  $P[\text{diagnosis} = \text{braindamaged} \mid \text{test} = \text{braindamaged}] \geq .80$  which can be crossvalidated by the leave-out-one (jackknife) technique of Mosteller & Tukey and Lachenbruch.

Whereas in classical diagnostics researchers are only trying to maximize hit rates we want to increase the *trustworthiness* of the prognosis, too. The trustworthiness increases with the critical distance of the prognosis to the cutoff. It can be shown that rules of prognosis with similar hit rates show large differences in their trustworthiness.

#### Literatur

Adams, J. (1969): On reconciling the 'multidimensional' and 'unitary' concepts of brain damage. *Percept. Motor Skills*, **29**, 579–598.

– (1970): Canter background interference procedure applied to diagnosis of brain damage in mentally retarded children. *Amer. J. Ment. Def.*, **75**, 57–64.

Aftanas, M. S. & Royce, J. R. (1969): A factor analysis of brain damage tests administered to normal subjects with factors scores. *Multivariate Behavioral Research*, **4**, 459–481.

Bauman, E. & St. John, J. (1971): The clinical usefulness of some tests of visual perception. *Psychology in Schools*, **8**, 247–249.

Bender, L. (1938): A visual motor Gestalt test and its clinical use. American Orthopsychiatric Association Research Monographs, No. 3, Menasha.

Birch, H. G. (1964): Summary of conference. *Am. J. Ment. Def.*, **64**, 410–420.

Boder, E. (1966): A neuropsychiatric approach in the diagnosis and management of school behavioral and learning disorders. In: Hellmuth, J., *Learning disorders*, Vol. II, Seattle, Washington, 1966.

Bräutigam, W. (1969): *Reaktionen, Neurosen, Psychopathien*. Stuttgart.

Brillinger, D. R. (1964): The asymptotic behavior of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Review of the International Statistical Institute*, **32**, 202–206.

Burgess, M. M., Kodanaz, A. & Ziegler, D. (1970): Prediction of brain damage in a neurological population with cerebrovascular accidents. *Percept. Mot. Skills*, **31**, 595–601.

– & Greenburg, H. (1970): Prediction of brain damage in two clinical populations. *Percept. Mot. Skills*, **30**, 523–532.

Canter, A. (1966): A background interference procedure to increase the sensitivity of the Bender-Gestalt Test to organic brain damage. *J. Consulting Psychology*, **30**, 91–97.

– (1968): The BIP-Bender test for the detection of organic brain disorder: modified scoring method and replication. *J. Consulting Clinical Psychology*, **32**, 522–526.

– (1970): The Canter Background Interference Procedure for the Bender-Gestalt test. Manual for administration, scoring, and interpretation. IOWA Psychopathic Hospital, Iowa City, Iowa.

– (1976): The Canter Background Interference Procedure. Manual. Nashville, Tennessee.

Campbell, S. B., Douglas, V. I. & Morgenstern, G. (1971): Cognitive styles in hyperactive children and the effect of methylphenidate. *J. Child. Psychol. Psychiat.*, 12, 55–67.

Clements, S. D. (1966): Minimal brain dysfunction in children. U.S. Department of Health, Education and Welfare, Washington, D.C.

– & Peters, J. E. (1965): The diagnostic dilemma in the child guidance. In: Hellmuth, J., *Learning disorders*, Vol. I, 197–218. Seattle, Washington.

Cohen, N. C., Weiss, G. & Minde, K. (1972): Cognitive styles in adolescents previously diagnosed as hyperactive. *J. Child Psychol. Psychiat.*, 13, 203–209.

Coleman, J. S. (1971): The mathematical study of change. In: Blalock, H. M. & Blalock, A. B. (eds.), *Methodology in Social Research*, N.Y., 428–478.

Cronbach, L. J. (1970): *Essentials of psychological testing*. N.Y.

Cureton, E. E. (1951): Approximate linear restraints and best predictor weights. *Educational and Psychological Measurement*, 11, 12–15.

Dallmeyer, H. J. (1969): Verhaltensauffälligkeiten nach früherworbener Hirnschädigung bei Kindern aus einer Erziehungsberatungsstelle. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 18, 273–278.

Dempster, A. P. (1966): Estimation in multivariate analysis. In: Krishnaiah, A. P. (ed): *Multivariate Analysis*, N.Y., Vol. I, 315–334.

Duhm, E. (1969): Psychologische Befunde bei Kindern mit Hirnschäden. *Beiträge zur Erziehungsberatung*, Bd. 4, 61–80, Weinheim.

Finifter, B. M. (1972): The generation of confidence evaluating research findings by random subsample replication. In: Costner, H. L. (ed): *Sociological Methodology*, 1972, San Francisco.

Frank, R. E., Massy, W. F. & Morrison, D. G. (1965): Bias in multiple discriminant analysis. *Journal of Marketing Research*, 2, 250–258.

Frankel, M. R. (1971): *Inference from survey sample*. Institute for Social Research, Ann Arbor.

Frierson, E. C. & Barbe, W. B. (1967): *Educating children with learning disabilities*. N.Y.

Fukunuga, F. & Kessel, D. (1971): Estimation of classification error. *IEEE Transactions on Electronic Computers*, C-20, 1521–1527.

Göllnitz, G. (1961): *Die Bedeutung der frühkindlichen Hirnschädigung für die Kinderpsychiatrie*. Leipzig.

Graham, Ph. & Rutter, M. (1968): Organic brain dysfunction and child psychiatric disorder. *British Medical Journal*, 3, 695-700.

Hall, M., Talmadge, M. & Hayden, B. S. (1968): Diagnosing minimal brain damage in children: a comparison of two Bender scoring procedures. Paper delivered at the annual meeting of the Eastern Psychological Association, Washington, D.C.

Halstead, W. C. & Rennick, Ph. M. (1966): Perceptual-cognitive disorders in children. In: Kidd, A. H. & Rivoire, J. L.: *Perceptual development in children*. N.Y., 5-31.

Harper, A. E., Jr. (1950): Discrimination between matched schizophrenics and normals by the Wechsler-Bellevue scale. *J. Consulting Psychology*, 14, 351-357.

Haynes, J. R. & Sells, S. B. (1963): Assessment of organic brain damage by psychological tests. *Psychological Bulletin*, 60, 316-325.

Herzberg, P. A. (1969): The parameters of cross-validation. *Psychometrika Monograph Supplement*, No. 2.

Hills, M. (1966): Allocation rules and their error rates. *Journal of the Royal Statistical Society*, B 28, 1.

Jensen, A. R. (1972): Hierarchical theories of mental ability. In: Jensen, A. R.: *Educational Differences*. London, 19-85.

Johnson, D. J. & Myklebust, H. R. (1968): *Learning disabilities*. New York.

Kanfer, F. H. & Saslow, G. (1974): *Verhaltenstheoretische Diagnostik*. In: Schulte, D.: *Diagnostik in der Verhaltenstherapie*. München, 24-29.

Kaspar, J. C. & Schulman, J. L. (1972): Organic mental disorders: brain damages. In: Wolman, B.: *Manual of child psychopathology*, 207-229.

Kenny, T. J. (1971): Background interference procedure: a means of assessing neurologic dysfunction in school aged children. *Journal of Consulting Psychology*, 37, 44-46.

Keogh, B. K. & Smith, C. E. (1968): Changes in the copying ability of young children. *Percept. Motor Skills*, 26, 773-774.

Kerekjarto, M. v. (1963): Untersuchungen über die Diskriminierungskraft dreier Tests zur Erfassung zerebraler Schäden. In: Lienert, G. A. (Hrsg.): *Bericht über den 23. Kongress der deutschen Gesellschaft für Psychologie*. Göttingen, 186-187.

Klee, E. (1974): *Behinderten Report*. Frankfurt.

Koppitz, E. M. (1964): *The Bender Gestalt test for young children*. N.Y.

Kornmann, R. (1971): *Hirnschädigung und fehlende Schulreife*. Berlin.

Krop, H. & Cohen, E. (1972): Predictive validity of the BIP Bender test. Paper presented at the meeting of the South Eastern Psychological Association. Atlanta.

Kuhlen, V. (1972): *Verhaltenstherapie im Kindesalter*. München.

Lachenbruch, P. A. (1965): Estimation of error rates in discriminant analysis. Ph. D. dissert., University of California at Los Angeles.

- Lachenbruch, P. A. (1975): *Discriminant Analysis*. N.Y.
- Lempp, R. (1964): *Frühkindliche Hirnschädigung und Neurose*. Bern.
- (1967): *Eine Pathologie der psychischen Entwicklung*. Bern:
- (1971a): *Lernerfolg und Schulversagen*. München.
- (1971b): *Organische Psychosyndrome*. In: Harbauer, H., Lempp, R., Nissen, G. & Strunk, P: *Lehrbuch der speziellen Kinder- und Jugendpsychiatrie*. Berlin, 209–241.
- Lord, F. M. (1963): *Elementary models for measuring change*. In: Harris, Ch. W. (ed): *Problems in measuring change*, 21–38.
- Lynn, R. (1960): *Individual differences in introversion-extraversion, reactive inhibition, and reading attainment*. *Journ. educ. Psychol.*, **51**, 318–321.
- Masini, A. (1969): *The background interference procedure and the Bender Gestalt test in the discernment of organic brain damage*. Unpublished Master's Thesis. Louisiana State University.
- Mecke, V. (1969): *Centration: a perceptual process diacritic of intellection and a differential diagnostic criterion*. *Percept. Motor Skills*, **29**, 827–834.
- Miller, R. G. (1964): *A trustworthy jackknife*. *Annals of Mathematical Statistics*, **34**, 567–582.
- Morrison, D. G. (1969): *On the interpretation of discriminant analysis*. *Journal of Marketing Research*, **6**, 156–163.
- Mosier, C. I. (1951): *Problems and designs of cross-validation*. *Educational and Psychological Measurement*, **11**, 5–11.
- Mosteller, F. & Tukey, J. W. (1968): *Data analysis, including statistics*. In: Lindzey, G. & Aronson, E. (eds): *Handbook of Social Psychology*, Vol. II, 80–203, Reading, Mass.
- Müller, D. (1967): *Diagnostische und psychopathologische Aspekte bei Kindern mit frühkindlichen zerebralen Schädigungen*. In: Szewczyk, H. & Rösler, H. D. (Hrsg.): *Probleme der klinisch-psychologischen Diagnostik*. Berlin.
- Müller-Küppers, M. (1969): *Das leicht hirngeschädigte Kind*.
- Mumenthaler, M. (1971): *Neurologie*. Stuttgart.
- Myklebust, H. & Boshes, P. (1967): *Psychoneurological learning disorders in children*. In: Frierson, E. C. & Barbe, W. B. (eds): *Educating children with learning disabilities*. N.Y., 16–35.
- Nicholson, G. E. Jr. (1960): *Prediction in future samples*. In: Olkin, I., Ghurye, S. G., Hoefding, W., Madow, W. G., Mann, H. B. (eds): *Contributions to Probability and Statistics*. Stanford, Calif.
- Paine, R. S. (1965): *Organic neurological factors related to learning problems*. In: Hellmuth, J. (Hrsg.): *Learning disorders*, Vol. I, 1–29. Seattle, Washington.

Pichot, P. & Persé, J. (1952): L'application des fonctions discriminantes en diagnostic individuel en psychologie. *Rev. Psychol. Appl.*, 2, 19-34.

Prüssing, O. (1967): Die Projektionstachistoskopie und ihre diagnostischen Möglichkeiten bei Hirngeschädigten. In: Szewczyk, H. & Rösler, H. D. (Hrsg.): *Probleme der klinisch-psychologischen Diagnostik*. Berlin, 173-184.

Quenouille, M. H. (1956): Notes on bias in estimation. *Biometrika*, 43, 353-360.

Reitan, R. M. (1962): Psychological deficit. *Annual Review of Psychology*, 13, 415-444.

Reitan, R. M. & Heineman, Ch. E. (1968): Interactions of neurological deficits and emotional disturbances in children with learning disorders: methods of differential assessment. In: Hellmuth, J. (Hrsg.): *Learning disorders*, Vol. III, 93-135. Seattle, Washington.

Schlange, H., Stein, B., Boetticher, I. v. & Taneli, S. (1972): Der Göttinger Formreproduktionstest (G-F-T). Zur Diagnose der Hirnschädigung im Kindesalter. Beiheft Handanweisung. Göttingen.

Schneeweiss, H. (1971): *Ökonometrie*. Würzburg.

Scholtz, W. (1972): *Testpsychologische Untersuchungen bei hirngeschädigten Kindern*. Berlin.

Schulte, D. (1974): *Diagnostik in der Verhaltenstherapie*. München.

Schulte, W. & Tölle, R. (1971): *Psychiatrie*. Berlin.

Smirnow, W. I. (1966): *Lehrgang der höheren Mathematik, II*. Berlin.

Specht, F. (1969): Fehlentwicklungen nach früherworbenen Hirnschäden. In: *Beiträge zur Erziehungsberatung*. Bd. 4, 83-98. Weinheim.

Stein, K. (1961): Some effects of brain damage upon speed, accuracy, and improvement in visual motor functioning. *J. Consulting Psychology*, 25, 171-178.

Tarnopol, L. (1971): *Learning disorders in children: diagnosis, medication and education*. Boston.

Thorndike, R. M., Weiss, D. J. & Davis, R. V. (1968): Multivariate relationship between a measure of vocational needs. *Journal of Applied Psychology*, 52, 491-496.

Tukey, J. W. (1958): Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, 29, 614.

Wallasch, R. & Möbus, C. (1977): Validierung und Kreuzvalidierung des Göttinger Formreproduktionstests von Schlange et al. (1972) und der Background Interference Procedure von Canter (1970) zur Erfassung von Hirnschädigungen bei Kindern zusammen mit zwei anderen Auswertungssystemen für den Bender Gestalt Test sowie weiteren Verfahren. *Diagnostica*, 2, 156-172.

Webster, H. & Bereiter, C. (1963): The reliability of changes measured by mental test scores. In: Harris, Ch. W. (ed): *Problems in measuring change*, 39-59. Madison.

Werry, J. S. (1968): The diagnosis, etiology and treatment of hyperactivity in children. In: Hellmuth, J. (ed): Learning disorders, Vol. III, 171–190. Seattle, Washington.

– (1972): Organic factors in childhood psychopathology. In: Quay, H. C. & Werry, J. S. (eds): Psychopathological disorders of childhood, 83–121. N.Y.

Wewetzer, K. H. (1956): Bender Gestalt Test bei Kindern: Auswertungsmethode und differentialdiagnostische Möglichkeiten. Zeitschrift für diagnostische Psychologie, 4, 174–186.

– (1959): Das hirngeschädigte Kind. Stuttgart.

Wheeler (1961): Symposion: An application of discriminant functions to the problem of predicting brain damage using behavioral variables. Paper presented at the Midwestern Psychological Association, Chicago.

– (1964): Complex behavioral indices weighted by linear discriminant functions for the prediction of cerebral damage. Percept. Motor Skills, 19, 907–923.

Wherry, R. (1951): Comparison of cross-validation with statistical inference of betas and multiple R from a single sample. Educational and Psychological Measurement, 11, 23–28.

Wing, J. K. (1973): Frühkindlicher Autismus. Weinheim.

Yates, A. J. (1954): The validity of some psychological tests of brain damage. Psychological Bulletin, 51, 358–379.

– (1966): Psychological deficit. Annual Review of Psychology, 17, 111–143.

Yulis, S. (1970): Performance of normal and organic brain damaged subjects on the Canter background interference procedure test as a function of drive. Journal Consulting Clinical Psychology, 34, 184–188.

Zimet, C. N. & Fishman, D. B. (1970): Psychological deficit in schizophrenia and brain damage. Annual Review of Psychology, 21, 113–154.

**Anschriften der Verfasser:**

Dr. Claus Möbus  
Psychologisches Institut  
der Universität Heidelberg  
Hauptstraße 47–51  
6900 Heidelberg

Dipl.-Psych. Roland Wallasch  
Psychologisches Institut  
der Universität Tübingen  
Friedrichstraße 21  
7400 Tübingen



## Anhang

### Anleitung zur Durchführung des „Bender Gestalt Tests“ und der „Background Interference Procedure“

Die Aufgabe des Patienten bei der Durchführung des „Bender Gestalt Tests“ und der „Background Interference Procedure“ besteht im Abzeichnen von neun, auf separaten Karten dargebotenen Stimulusfiguren. Die Karten werden jeweils einzeln und in der Reihenfolge A, 1, 2, ..., 8 vorgelegt. Dabei ist darauf zu achten, daß der Patient immer nur die gerade zum Abzeichnen vorgegebene Karte sieht.

Zum Zeichnen benötigt der Patient einen weder zu harten noch zu weichen Bleistift, am Besten 2B, der aber nicht so stark gespitzt sein sollte, daß er gleich abbricht.

Im ersten Durchgang („Bender Gestalt Test“) werden die Stimulusfiguren auf normales weißes Papier gezeichnet. Es ist darauf zu achten, daß das Papier *hochkant* vor dem Patienten liegt. Zudem wird der Patient angewiesen, das Papier beim Zeichnen nicht aus dieser Position zu drehen (s. unten, Instruktion).

Nach dem ersten Durchgang wird für einen Zeitraum von ca. 10 Minuten eine andere Tätigkeit interpoliert (z. B. Anamnese, Interview, andere Untersuchungen, Pause zum Entspannen, o. ä.). Die Aktivität in diesem Zeitraum sollte unbedingt für alle Patienten gleich sein, da evtl. damit zu rechnen ist, daß sie einen Einfluß auf das nachfolgende zweite Abzeichnen der Figuren hat und dieser Einfluß möglichst konstant gehalten werden sollte. Für den Patienten körperlich und/oder geistig anstrengende Aktivitäten sollten in diesem Zeitraum ebenfalls vermieden werden.

Nach dieser interpolierten Aufgabe werden im zweiten Durchgang („Background Interference Procedure“) dieselben Figuren erneut vom Patienten abgezeichnet, diesmal auf das Interferenz-Papier. Auch hier muß das Blatt hochkant liegen (bleiben). Versucht der Patient, das Papier während des Zeichnens zur Erleichterung der Aufgabe zu drehen, so bitten Sie ihn, das Papier hochkant liegen zu lassen und es auf jeden Fall so zu versuchen.

Die folgenden Instruktionen sollten wörtlich gegeben werden, um gleiche Bedingungen für alle Probanden zu gewährleisten.

#### Instruktion zum ersten Durchgang („Bender Gestalt Test“)

„Ich möchte Sie jetzt bitten, etwas abzuzeichnen. Ich habe hier neun Karten. Auf jeder Karte ist eine Figur gezeichnet, und ich lege Ihnen immer eine Figur vor und bitte Sie, diese so gut Sie können auf das vor Ihnen liegende Papier abzuzeichnen. Zeichnen Sie alle Figuren auf dieses eine Blatt. Das Papier muß so liegen bleiben wie es jetzt liegt, nämlich hochkant. Bitte drehen Sie es beim Zeichnen nicht in eine andere Lage, auch dann nicht, wenn es dann leichter ginge. Zeichnen Sie bitte freihändig ohne Lineal oder Hilfsmittel und ohne zu stricheln. Versuchen Sie, die Figur so groß zu zeichnen, wie sie auf der Karte ist.“

*Hinweis:* Wenn der Patient fragt, was „stricheln“ bedeutet, machen Sie es ihm bitte auf separatem Papier vor.

*Beispiel:*

Notieren Sie bitte alle wichtigen Auffälligkeiten bei der Durchführung des Tests, z. B. Versuche zu radieren, Linien mit dem Finger zu verwischen oder wegzukratzen.

Verlangt der Patient, eine mißlungene Figur ein zweites Mal zeichnen zu dürfen, so erlauben Sie es ihm. In diesem Fall kennzeichnen Sie bitte nach der Testdurchführung den ersten und zweiten Versuch. Mehr als zwei Zeichnungen einer Stimulusfigur sollen grundsätzlich nicht erlaubt werden.

Instruktion zur „Background Interference Procedur“  
(zweiter Durchgang)

Nach der interpolierten Aufgabe und dem Bleistiftspitzen legen Sie das Interferenz-Papier hochkant vor den Patienten und instruieren Sie ihn wie folgt:

„Jetzt zeichnen Sie bitte die Figuren, die Sie vorhin auf das weiße Papier gezeichnet haben, auf dieses Papier noch einmal ab. Machen Sie es bitte wie zuvor – zeichnen Sie jede Figur so groß, wie sie auf der Karte ist.“

Fragt der Patient, ob er die Figuren in die weißen Zwischenräume zwischen den Linien auf dem Interferenz-Papier zeichnen soll, so antworten Sie:

„Zeichnen Sie die Figuren bitte so ab, wie Sie es vorher gemacht haben. Zeichnen Sie sie in der gleichen Größe wie auf der Karte.“

Zeichnet der Patient trotzdem eine Figur in einen weißen Zwischenraum, so unterbrechen Sie ihn sobald er mit dieser Figur fertig ist und fordern Sie ihn erneut auf, „so groß wie vorher“ zu zeichnen.

Beharrt ein Patient trotzdem weiterhin auf diesem Vorgehen, so lassen Sie ihn gewähren und fordern Sie ihn nach Beendigung des gesamten Tests auf:

„So, jetzt möchte ich aber einmal sehen, ob Sie die Figuren so zeichnen können, daß sie genau so groß sind wie auf der Karte. Zeichnen Sie einfach über die Linien hinweg, als ob sie nicht da wären.“