

Aus der Psychiatrischen und Nervenkl. des Städt. Krankenhauses Lübeck
(Chefarzt: Prof. Dr. GERHARD SCHMIDT)

Der diagnostische Wert graphischer Tests in der medizinischen Psychologie

Von

RICHARD SUCHENWIRTH, VASIL FILIPIDIS und HEINZ KOTTENHOFF

(Eingegangen am 24. Dezember 1963/15. April 1964)

In den letzten Jahrzehnten ist mehrfach versucht worden, Zeichentests zu entwickeln und zu standardisieren, die auch für die Neuro-psychiatrie brauchbar sind.

All diese Verfahren erlauben auf die eine oder andere Weise Aussagen über die graphische Funktion der untersuchten Probanden. Rückschlüsse auf die klinische Diagnose können nur indirekt gezogen werden. So fehlt es nicht an kritischen und skeptischen Stimmen, die den Nutzen dieser Verfahren für die klinische Psychologie bezweifeln. Am weitesten ging dabei wohl BRENGELMANN, der jüngst zwar anerkannte, daß „eine signifikante Wechselwirkung zwischen Persönlichkeit und graphischem Ausdruck besteht“, zugleich aber feststellt, „daß die klinische Validierung des zeichnerischen Ausdrucks praktisch erfolglos geblieben“ ist. Er fügt hinzu: „Es ist bedauerlich, daß sie“ (die graphischen Verfahren) „für praktische Zwecke übernommen und empfohlen werden, solange ihre diagnostischen Qualitäten rein hypothetisch sind.“

Wer die Schwierigkeiten der Diagnosestellung in der Nervenheilkunde, überhaupt in der Medizin, kennt, wird allerdings kaum erwarten, daß eine einzelne Untersuchungsmethode für einen psychischen Teilbereich zur klinischen Diagnose führen kann. Fast immer ist in mehrdimensionaler Betrachtungsweise die Anwendung zahlreicher Untersuchungsmethoden notwendig.

I

Für die praktische Arbeit wäre schon viel gewonnen, wenn wir Verfahren besitzen, die sowohl Aussagen über einen untersuchten Funktionsbereich erlauben, als auch gewisse — mehr als zufällige — Rückschlüsse auf die klinische Diagnose zulassen.

Im Rahmen einer größeren Untersuchung haben wir einige geläufige graphische Testmethoden daraufhin geprüft, ob sie zu medizinisch-diagnostischen Aufschlüssen verhelfen, welches Verfahren dabei die besten Ergebnisse hat und bei welchen klinischen Hauptgruppen die Untersuchungen am aufschlußreichsten sind.

Da die Tests (angewandt wurden der *Wartegg-Zeichen-Test (WZT)*, der *Baumzeichentest (BZT)*, der *Bender-Gestalt-Test (BGT)* ein *Kritzeltest* nach KUTASH-GEHL und die Prüfung einer *Schriftprobe*) hier nur in einer Nebenfunktion überprüft werden, — keines der Verfahren ist an sich dafür gedacht, eine neuropsychiatrische Diagnose zu stellen — wird von der Nullhypothese ausgegangen; d. h. es wird angenommen, daß bei Blindauswertung der Testprotokolle nicht mehr, als rein zufällig zu erwarten, richtige Diagnosen gestellt werden.

II

Untersucht wurden 50 erwachsene Probanden mit dem WZT, BZT, BGT, Kritzel-Test und an Hand einer Schriftprobe, und zwar (je 5 Männer und 5 Frauen) je 10 Patienten mit der Diagnose Schizophrenie, Cyclothymie, Anfallsleiden und hirnorganische Erkrankung (diffuse Prozesse des Zentralnervensystems, z. B. multipler Sklerose), außerdem zehn gesunde Vergleichspersonen. Die Probanden waren nicht besonders ausgewählt, es wurde lediglich darauf geachtet, daß die Diagnose als gesichert galt und daß sie willens und fähig gewesen waren zu zeichnen. (Schwere organische Nervenleiden und Psychosen im akuten Stadium schieden dadurch aus). Die Schulbildung der Probanden in den einzelnen Gruppen war gleich. Die Probanden benötigten 30–75 min für die Lösung der Testaufgaben. Die Zeit wurde registriert, aber nicht weiter verwertet.

Die 250 ausgefüllten Blätter wurden verschlüsselt und dann willkürlich durcheinander gemischt sechs Auswertern mit unterschiedlicher klinischer und testpsychologischer Erfahrung vorgelegt. Die Auswerter wußten, um welche klinischen Gruppen es sich handelt und sollten relativ schnell intuitiv oder unter Berücksichtigung der nur einem Teil der Auswerter bekannten empirischen Kriterien (für den WZT nach MÜLLER-SUUR und PFEIFFER, für den BGT nach BENDER und KOTTENHOFF, für den Kritzeltest nach GEHL und KUTASH, für den BZT nach SUCHENWIRTH, MORITZEN u. HAUSS) die Protokolle den einzelnen Gruppen zuordnen. Die Untersucher benötigten je 2–4 Std, also pro Einzeltest nicht einmal 2 min.

Es wurde anschließend ausgezählt, wie viele Lösungen richtig waren. Dann wurde geprüft, ob die Zahl der richtigen Ergebnisse noch als zufällig angesehen werden konnte (Heranziehung der Binomialverteilung, 99⁰/₁₀₀-Bereich).

III

Die Einzelergebnisse und die Ergebnisse in den Gruppen müssen den beiliegenden Tabellen entnommen werden.

BZT, BGT und WZT hatten zu signifikant mehr richtigen Lösungen geführt, als rein zufällig zu erwarten war. In der Normalgruppe, bei den

Hirnorganisch-Kranken, den Anfallspatienten und den Schizophrenen waren die Ergebnisse gleichfalls signifikant besser, als nach der Nullhypothese zu erwarten. Die Gruppe der Cyclothymen wurde von allen Untersuchern der Nullhypothese entsprechend identifiziert.

Die ungeübten Auswerter (D: keine klinische, keine test-psychologische Erfahrung und A) hatten die schlechtesten Ergebnisse, die geübteren Auswerter (B, C, E, F: durchschnittlich 6 Jahre klinische Erfahrung, Erfahrung mit den graphischen Testmethoden) deutlich mehr Erfolge. Die Erfahrungen von SCHMIDT u. McGOWAN beim Figurzeichentest, daß ungeübte und geübte Auswerter ähnliche Ergebnisse haben können, bestätigten sich bei uns nicht.

IV

Diese Ergebnisse lassen folgende Deutung zu:

Die graphischen Testmethoden erlauben über ihren eigentlichen (vorwiegend charakterologischen) Einsatzbereich hinaus gewisse statistisch erhärtbare Rückschlüsse auch auf die klinische Diagnose. Dies gilt vor allem bei Hirnorganisch-Kranken, weniger bei Anfallskranken und Schizophrenen. Bei Cyclothymen ist diese Anwendung der Tests sinnlos.

Es ist hervorzuheben, daß der Zeitaufwand für den Untersucher äußerst gering ist. Übung und Beschäftigung mit den theoretischen Grundlagen dieser Verfahren sind auch bei dieser Anwendungsweise notwendig.

Das Ergebnis ist beachtlich, wenn man bedenkt, daß die klinischen Gruppen durch die heute angewandte Therapie (Psychopharmaka, Elektroschocks) sich nicht mehr ganz so charakteristisch voneinander unterscheiden wie früher. Auch muß bedacht werden, daß die Gruppe der Hirnorganisch-Kranken fließende Übergänge zu den Anfallskranken zeigt.

Andererseits bleibt der diagnostische Wert — wie gar nicht anders zu erwarten und in der geringen Trefferzahl deutlich — recht beschränkt, wenn die Frage der klinischen Diagnose gestellt wird.

V

Psychometrische Schlußbemerkungen (von H. K.)

Gemäß SUCHENWIRTH et al. wäre die Erwartungshypothese für die Zufallstreffer während des diagnostischen Klassifikationsvorganges als konstant zu betrachten. Die Konstanzannahme gilt in der Psychometrie unter der Voraussetzungsalternative, daß

a) die diagnostisierten Fälle in die Stichproburne zurückgelegt werden, bevor die nächste Falldiagnose vorgenommen wird; in diesem Falle ist nämlich die Stichprobengröße und -zusammensetzung (aus verschiedenen Fällen) die gleiche für jede einzelne diagnostische Entscheidung.

b) die Stichprobenpopulation als sehr groß (unendlich groß) betrachtet werden kann; untersuchen wir z. B. (Beispiel nach YULE u. KENDALL 1950, p. 395) eine Bevölkerung, die 23 Millionen Frauen und 22 Millionen Männer enthält, mit einer Stichprobe von 1000 Personen. Die Chance, beim ersten Griff in die Urne einen Mann zu erwischen, beträgt

$${}_1P = 22\,000\,000 / 45\,000\,000.$$

Auch nach Untersuchung von 999 Fällen ist die Chance beim tausendsten Griff immer noch

$${}_{1000}P = 21\,999\,001 / 44\,999\,001.$$

Beide P, nämlich das beim ersten und das beim 1000. Griff, können als annähernd gleich groß aufgefaßt werden, solange die Population groß ist.

Nachdem Voraussetzungsalternative (a) in der obigen diagnostischen Untersuchung nicht vorlag, bleibt zu fragen, ob (b) zutraf, bzw. für welche diagnostischen Kategorien es zutraf. Es wäre ja denkbar, daß die (klinische) Population, denen SUCHENWIRTH et al. ihre fünf verschiedenen Stichprobenkategorien (Schizophrene, Cyclothyme, Anfallsleiden und Hirnorganiker sowie eine Vergleichsstichprobe normaler Versuchspersonen) entnahmen, ursprünglich gar nicht in diese fünf Gruppen gleichmäßig zerfällt, sondern daß z. B. einer Vielzahl von Schizophrenen und Hirnorganikern nur eine verschwindend kleine Zahl von Cyclothymen und Anfallsleiden gegenübersteht, so daß von SUCHENWIRTH et al. als Stichprobenfälle der letzteren beiden Kategorien gewählten Personen das Gesamtkontingent der Klinik (deren „Population“) nahezu erschöpfen. In diesem Falle würde es für diese letzteren Kategorien (b) nicht mehr gelten, und wir müssen die Berechnung der Zufallstreffer nach anderen Regeln vornehmen. (Der Fall ist nur als Beispiel konstruiert. Tatsächlich ist die Stichprobenpopulation bei diesen Hauptgruppen psychiatrischer Erkrankungen groß.)

Diese Regeln sind allerdings ohne höhere Mathematik, wie sie z. B. WITTES „Einführung in die mathematische Behandlung psychologischer Probleme“ anwendet, nur schlecht verstehbar. Deshalb wollen wir hier auf eine Exemplifizierung verzichten und statt dessen die Logik der Zufallstrefferberechnung und Signifikanzbestimmung, wie sie SUCHENWIRTH et al. im obigen Artikel, Tafeln 1, 2 und 3 unter Berufung auf die „Geigytabellen“ ausgeführt haben, etwas näher betrachten, wobei wir die Stichprobenpopulation als unendlich groß (die gestellten Diagnosen also als allgemein repräsentativ für die betreffenden Krankheiten) voraussetzen.

Bei den Tab. 1, 2 und 3 handelt es sich um Häufigkeitsziffern, denen die Annahme einer Binomialverteilung zugrunde gelegt wird. Um den 99%igen Mutungsbereich der Binomialverteilung zu ermitteln, haben

Tabelle 1

Richtige Lösungen bei den einzelnen Auswertern und zu erwartende zufällige Lösungen

Auswerter A:

Gruppe	WZT	BZT	BGT	Kritzeln	Schrift	zus.	rein zufällig zu erwarten
Schizo.	1	4	0	2	0	7	10 (3,9—19)
Cycloth.	3	0	2	2	1	8	10 (3,9—19)
Hirnerg.	2	5	3	6	5	21	10 (3,9—19)
Anfalls.	2	1	1	1	2	7	10 (3,9—19)
Normal	5	4	5	3	4	21	10 (3,9—19)
	13	14	11	14	12	64	50 (34,9—67,9)

Auswerter B:

Schizo.	2	5	3	keine Ergebnisse	3	13	8 (2,7—16)
Cycloth.	2	2	1		1	6	8 (2,7—16)
Hirnerg.	1	6	1		3	12	8 (2,7—16)
Anfalls.	1	4	3		4	11	8 (2,7—16)
Normal	5	3	6		6	20	8 (2,7—16)
	11	20	14		17	62	40 (26,4—56,36)

Auswerter C:

Schizo.	3	3	2	keine Ergebnisse	8	6 (1,7—13,2)
Cycloth.	2	2	—	keine Ergebnisse	4	6 (1,7—13,2)
Hirnerg.	3	3	3	keine Ergebnisse	9	6 (1,7—13,2)
Anfalls.	3	3	2	keine Ergebnisse	8	6 (1,7—13,2)
Normal	6	4	4	keine Ergebnisse	14	6 (1,7—13,2)
	17	15	11		43	30 (18,37—44,4)

Auswerter D:

Schizo.	1	2	2	3	2	10	10 (3,9—19)
Cycloth.	3	2	2	0	1	8	10 (3,9—19)
Hirnerg.	3	3	3	3	3	15	10 (3,9—19)
Anfalls.	0	4	6	0	3	13	10 (3,9—19)
Normal	0	2	1	4	5	12	10 (3,9—19)
	7	13	14	10	14	58	50 (34,9—67,9)

Auswerter E:

Schizo.	5	9	2	2	3	21	10 (3,9—19)
Cycloth.	2	4	3	3	1	13	10 (3,9—19)
Hirnerg.	3	6	4	3	2	18	10 (3,9—19)
Anfalls.	4	6	3	1	4	18	10 (3,9—19)
Normal	2	5	5	5	4	21	10 (3,9—19)
	16	30	17	14	14	91	50 (34,9—67,9)

Auswerter F:

Schizo.	4	3	2	1	3	13	10 (3,9—19)
Cycloth.	3	1	1	1	1	7	10 (3,9—19)
Hirnerg.	2	2	5	2	4	15	10 (3,9—19)
Anfalls.	4	5	1	1	2	13	10 (3,9—19)
Normal	6	5	7	3	5	26	10 (3,9—19)
	19	16	16	8	15	74	50 (34,9—67,9)

Tabelle 2. Vergleich der Ergebnisse mit den einzelnen Tests mit den rein zufällig zu erwartenden Ergebnissen
(Binomialverteilung, 99% Wahrscheinl.)

	WZT	BZT	BGT	Kritzelt.	Schriftpr.
Insges. ausgewert.:	300	300	300	200	250
Richtige Lösungen:	83	108	83	42	72
Prozent:	27,6%	36%	27,6%	21%	28,8%
Zu erwarten als richtig rein zufällig:	60	60	60	40	50
	(43,11—79,65)			(26,4—56,36)	(34,9—67,9)
Zu erwarten in Prozent an richt. Lösg.	14,37—26,55%			13,2—28,18%	13,87—27,23%

Tabelle 3. Ergebnis des WZT, BZT und BGT bei den klinischen Gruppen im Vergleich mit den als zufällig zu erwartenden Ergebnissen

	Schizo.	Cycloth.	Hirnerg.	Epil.	Normal
Insgesamt ausgewertet:	180	180	180	180	180
Richtige Lösungen:	53	35	58	53	75
Prozent:	29,4%	19,4%	31,5%	29,4%	41,6%
Zufällig als richtig zu erwarten:	36	36	36	36	36
	12,87—28,67%				

Die statistischen Tabellenangaben entstammen den „Wissenschaftlichen Tabellen“ Geigy (1962).

SUCHENWIRTH et al. die „Wissenschaftlichen Tabellen, J. R. Geigy A. B., Basel“ in der neuesten Ausgabe bemüht. Eine für den Psychometriker ebenso brauchbare und ausführliche Referenz ist E. WEBERS „Grundriß der biologischen Statistik“, Jena 1961, Tafel 8 (Neue prozentuale Konfidenzgrenzen); das folgende „Bienenbeispiel“ WEBERS klärt die Terminologie ab: Ein Bienenforscher untersucht aus einem Volk $n = 16$ Bienen; von diesen leiden $z = 5$ an Milbenkrankheit. Gefragt wird nach dem Anteil kranker Bienen im ganzen Volk:

$$\text{Mittelwert } p_z = (5) (100)/16 = 31,25 (\%)$$

Die Vertrauensgrenze dieses Mittelwertes kann man aus der Weberischen Tafel 8 wie folgt entnehmen. Man geht mit $n = 16$ und $z = 5$ in die Tafel ein und findet für

$$\begin{aligned} &\text{die untere Grenze } p_z = 8,8 (\%) \\ p_z &= 31,25 \text{ die obere Grenze } p_z = 64,0 (\%) \end{aligned}$$

Diese Grenzen umfassen 99% der Binomialverteilung, d. h. daß der Mittelwert p_z mit 99% statistischer Sicherheit innerhalb dieser Grenzen

zu suchen ist. Die Webersche Tafel 8 ist mit „alpha = 1%“ überschrieben; eine Tafel 8a, die mit „alpha = 5%“ betitelt ist, sagt uns, daß die 95%-Konfidenzgrenzen für das jetzige Bienenbeispiel nach unten bei 14,8(0/0), nach oben bei 56,7(0/0) liegen. Je schärfer, (d. h. kleiner) also alpha gewählt wird, um so größer wird das Konfidenzintervall ausfallen, außerhalb dessen man z. B. das Auftreten der Milbenkrankheit in einem Volk als von dem obigen Volk abweichend betrachten kann.

Wir ziehen nun die Suchenwirthsche Tab.1 in etwas modifizierter Form zur Demonstration heran; und zwar greifen wir die Resultate des Auswerter C heraus, weil dieser nur drei Testverfahren (WZT, BZT und BGT) benutzte, wodurch wir an Übersicht gewinnen. Die Häufigkeitszahlen schreiben wir besser als Proportionszahlen auf; es waren pro Versuchsgruppe (Diagnose) je 10 Fälle untersucht worden, was bei fünf Versuchsgruppen gleicher Größe 50 Fälle ergibt:

Diagnosegruppe	WZT	BZT	BGT	zusammen (od.) in Prozent	
					%
Schizo.	3/10	3/10	2/10	8/30	26,6
Cyclo.	2/10	2/10	0/10	4/30	13,3
Hirnorg.	3/10	3/10	3/10	9/30	30,0
Epil.	3/10	3/10	2/10	8/30	26,6
Normal	6/10	4/10	4/10	14/30	46,6
Zusammen	17/50	15/50	11/50	43/150	28,6

SUCHENWIRTH et al. geben außerdem hinter jeder (diagnostischen) Reihe der Tab.1 die nach obiger Zufallshypothese zu erwartenden Treffer an, z. B. für jede Reihe des obigen Ausschnittes 6/30. (Bei anderen Auswertern als C sind die zu erwartenden Zufallstreffer größer, weil diese mehr Verfahren anwandten). $6/30 = p_0$, nämlich p_z unter der Nullhypothese (H_0), daß die von dem Auswerter C erreichten diagnostischen Treffer von diesem Mittelwert nur zufallsmäßig abweichen, wobei eine Binomialverteilung zugrundegelegt und mit alpha als 1% festgelegt wurde. Wo liegen die 99%-Konfidenzgrenzen für den Mittelwert p_0 oder p_z ? Analog unserem Bienenbeispiel erhalten wir für $n = 30$ und $z = 6$ aus Webers Tafel

die untere Grenze $p_z = 6,2(0/0)$

$p_z = 20,0$

die obere Grenze $p_z = 43,0(0/0)$.

Drücken wir diese Grenzprozente in Häufigkeitsziffern aus, so erhalten wir:

untere (6,2) (30)/100 = 1,9

$f_z = 6$

obere (43,0) (30)/100 = 13,0.

Diese Zahlen entsprechen (annähernd, weil die Geigy-Tabellen offenbar noch eine besondere Korrektur einschließen) den Werten in Suchenwirths Tab. 1, Auswerter C, am rechten Rand.

Bemerkt sei noch, daß wir, nachdem wir uns einmal für das 99%ige Konfidenzintervall entschieden haben, nicht nachträglich zum 95%igen Konfidenzbereich referieren dürfen. Unter strenger Einhaltung dieser Regel würde also der Auswerter C nur die Normalpersonen als „überzufällig richtig“ mittels der drei Testverfahren einstufen, nicht dagegen die Kranken (Schiz., Cyclo. etc.). Dies erklärt uns die Unterstreichung der Häufigkeitsziffer 14 in der Reihe der Normal-Versuchspersonen.

Zusammenfassung

Geprüft wurde, ob die geläufigsten graphischen Testmethoden über ihren eigentlichen Anwendungsbereich hinaus mehr als zufällige Rückschlüsse auf die klinische Diagnose erlauben. Untersucht wurden 50 Probanden (je 10 Schizophrene, Cyclothyme, Hirnorganisch-Kranke, Anfalls-kranke und Gesunde) mit dem Wartegg-Zeichen-Test, dem Baumzeichen-Test, dem Bender-Gestalt-Test, einem Kritzeltest (nach GEHL-KUTASH) und an Hand einer kurzen Schriftprobe. Die Protokolle wurden von sechs unterschiedlich vorgebildeten Auswertern ohne Kenntnis der Kranken den klinischen Gruppen zugeordnet. BZT, WZT und BGT ergaben bei der Normalgruppe, den Anfallskranken, den Hirnorganisch-Kranken und den Schizophrenen signifikant etwas höhere Ergebnisse, als rein zufällig zu erwarten war. Übung und Erfahrung mit den Tests führten zu deutlich besseren Erfolgen. Die erwähnten Verfahren besitzen also über ihr eigentliches Anwendungsgebiet hinaus einen gewissen, wenn auch nicht großen Wert als diagnostisches Hilfsmittel, die Nullhypothese kann als widerlegt gelten. Die Trefferzahl dieser ausgesprochenen Schnell- und Kurzverfahren ist allerdings keineswegs so hoch — was auch gar nicht zu erwarten war —, daß sie mehr als die Stellung einer Hilfsuntersuchung verdienen.

Literatur

- BENDER, L.: A visual motor Gestalt-Test and its clinical use. Res. Monograph. Nr. 3, American Orthopsych. Assoz. New York 1938.
 BRENGELMANN, J. C.: Psychologische Methodik und Psychiatrie. Psychiatrie der Gegenwart I/2, S. 134—175. Berlin, Göttingen, Heidelberg: Springer 1963.
 KUTASH, S. B., and R. H. GEHL: The Graphomotor Projection Technique. Springfield, Ill. 1954.
 KOTTENHOFF, H.: Bender's Gestalt-Test und die Untersuchung des cerebralen Funktionswandels. Psychol. Beitr. 2, 40—55 (1955).
 MÜLLER-SUUR, H.: Psychiatrische Erfahrungen mit dem Wartegg-Zeichentest. Nervenarzt 23, 446—450 (1952).
 PREIFFER, W.: Der Zeichentest nach WARTEGG in der psychiatrischen Diagnostik (unter besonderer Berücksichtigung der Reizverarbeitung). Arch. Psychiat. Nervenkr. 187, 268—290 (1951).

- ROST, H.: Wartegg-Zeichen-Test-Untersuchungen bei Schizophrenen. Arch. Psychiat. Nervenkr. **189**, 251—278 (1952).
- SCHMIDT, L. D., and J. F. MCGOWAN: The differentiation of human figure drawings. J. cons. Psychol. **23**, 129—133 (1959).
- SUCHENWIRTH, R., u. J. MORITZEN: Untersuchungen mit dem Baumtest (KOCH) bei zyklotyphen und schizophrenen Kranken. Z. Psychother. med. Psychol. **10**, 52—63 (1960).
- , u. K. HAUSS: Der Baumzeichentest (KOCH) bei Anfallskranken. Z. Psychother. med. Psychol. **13**, 195—200 (1963).
- WITTE, W.: Einführung in die mathematische Behandlung psychologischer Probleme. In DORSCH: Psycholog. Wörterbuch, 7. Aufl. Hamburg u. Bern 1963.
- YULE, U., and M. G. KENDALL: An introduction to the theory of statistics, 14. Aufl. London 1950.

Dr. R. SUCHENWIRTH u. Dr. V. FILIPIDIS,

Lübeck, Ratzeburger Allee 160

Dr. H. KOTTENHOFF, psykolog

Karlstad/Schweden, Centrallasarettet Psykiatriska kliniken, Psykologiska
laboratoriet