

Methodologische Probleme bei der Effizienzprüfung von Psychopharmaka *

Hans Heimann

Psychiatrische Universitätsklinik Tübingen (Direktor: Prof. Dr. H. Heimann)

Eingegangen am 2. Juni 1975

Methodological Problems of Clinical Trials of Psychotropic Drugs

Summary. Up to now, clinical studies only succeeded in differentiating between great categories of psychotropic drugs, but failed to prove finer differences of effects within these categories of substances. Two points of the testing-method are discussed: 1. problems which arise when rating pathological behaviour and 2. problems of sampling psychiatric patients.

A great part of symptoms that clinicians and psychologists used to consider as relevant proved to be extremely rare. Total scores cannot be taken as a measure of the therapeutic effect, because they don't express adequately the degree of severity of the illness before and after treatment, and there is a parameter of symptom that is independant from the observer, i.e. the frequency of the symptom in different clinical pictures. The frequency is an inverse ratio to the specificity of the symptom.

It is then argued that even in clinical studies, it would be possible to choose among the variety of descriptive symptoms those which fulfil requirements of the probabilistic test-model of Rasch and to take only those symptoms to characterize the degree of severity of psychic disturbance in trials with psychotropic drugs.

Conclusions are then drawn from a study including three groups of physicians (specialists for internal diseases, psychiatrists and general practitioners): failure to differentiate between placebo and a Minor-Tranquillizer was not due to the inefficiency of the drug, but ought to be attributed to the lack of sharpness of the observations made by untrained judges. A significant difference between placebo and the Minor-Tranquillizer was yet found, but only in the group of psychiatrists. The comparison of the first 13 and the last 13 cases in the two remaining groups, however, reveals a learning process in the course of the study.

The main problems of sampling are discussed, i.e.: the loss of information as a consequence of taking the mean in a group of psychiatric patients, the role of biological rhythms, which was hitherto insufficiently considered, and finally it is demonstrated in connection with two selected cases of depressive patients that enormous difference of psychophysiological responsiveness can be hidden behind very similar clinical pictures.

It is pointed out that the existing research strategy is adjusted to great samples, which were composed on the basis of behavioral characteristics, and that it failed to differentiate subtle effects of psychotropic drugs. Only experiments involving a much greater display, which take into account all aspects of observation of the selected single cases and longitudinal studies can answer the question which is the right medicine for the right patient. Psychophysiological and biochemical methods have here priority over other methods.

Key words: Psychotropic Drugs — Testing-Methods — Psychophysiological Responsiveness — Rating Scales — Sampling Problems.

Zusammenfassung. Ausgehend von der Tatsache, daß es im klinischen Bereich nur gelungen ist, eindeutige Unterschiede zwischen den großen Kategorien von Psychopharmaka nachzuweisen, nicht jedoch feinere Wirkungsdifferenzen innerhalb dieser Präparatkategorien, werden zwei Problemkreise der Prüfungsmethodik diskutiert, 1. die Probleme, welche sich bei

* Nach einem Vortrag, gehalten auf dem Kongreß der Deutschen Gesellschaft für Psychiatrie und Nervenheilkunde, München, Oktober 1974.

der Merkmalserhebung von pathologischem Verhalten ergeben, und 2. Probleme der Stichprobenbildung bei psychiatrischen Patienten.

Es wird gezeigt, daß unter den zur Quantifizierung benutzten Merkmalen, welche Kliniker und Psychologen für relevant halten, eine große Zahl extrem selten vorkommt, daß die Gesamtscores zur Messung des therapeutischen Effektes den Schweregrad eines Krankheitsbildes vor und nach der Behandlung nicht adäquat ausdrücken und daß es einen Merkmalsparameter gibt, der vom Beobachter unabhängig ist, nämlich die Häufigkeit des Merkmals in verschiedenen klinischen Zustandsbildern. Sie verhält sich umgekehrt zur Spezifität der Merkmale.

Ferner wird dargestellt, daß es auch im klinischen Bereich möglich wäre, aus der Fülle deskriptiver Merkmale diejenigen auszuwählen, die den Anforderungen des probabilistischen Testmodells von Rasch entsprechen, und nur diese Merkmale zur Charakterisierung des Schweregrades psychischer Störungen in Psychopharmakoprüfungen heranzuziehen.

An den Resultaten eines Feldversuchs mit Internisten, Allgemeinpraktikern und Nervenärzten wird gezeigt, daß der fehlende Nachweis eines Unterschiedes zwischen Placebo und einem Minortranquillizer nicht auf der Unwirksamkeit des letzteren beruhte, sondern auf der ungenügenden Schärfe der Beobachtungen bei nicht geübten Beobachtern. Bei den Nervenärzten ergab sich ein signifikanter Unterschied zugunsten des Aktivpräparates, nicht jedoch bei den Allgemeinpraktikern und Internisten, die eine extrem hohe Placeborate aufwiesen. Wenn man jedoch die ersten 13 und die letzten 13 Fälle, die von Allgemeinpraktikern und Internisten behandelt wurden, vergleicht, ließ sich für diese Untersucher ein Lernprozeß während der Untersuchung nachweisen.

Unter den Problemen der Stichprobenbildung werden vor allem der Informationsverlust diskutiert, der bei psychiatrischen Patienten durch Mittelwertsbildung entsteht, die Frage der biologischen Rhythmen, die bisher nur ungenügend berücksichtigt wurden, und schließlich wird anhand von zwei ausgewählten Fällen psychophysiologischer Untersuchungen an depressiven Patienten demonstriert, welche riesigen Unterschiede in der psychophysiologischen Reaktivität hinter ähnlichen klinischen Bildern verborgen sein können.

Es wird darauf hingewiesen, daß die bisherige Forschungsstrategie, die vor allem auf große Stichproben ausgerichtet ist, welche auf Grund von Verhaltensmerkmalen zusammengestellt werden, die in sie gestellten Erwartungen einer Differenzierung feiner Psychopharmakawirkungen nicht erfüllen kann, sondern daß erst eine mit viel größerem Aufwand alle Beobachtungsbereiche berücksichtigende Untersuchung von ausgewählten Einzelfällen im Längsschnitt die Frage beantworten kann, welches das richtige Präparat für den richtigen Patienten sei. Dabei kommen den psychophysiologischen und den biochemischen Methoden vorrangige Bedeutung zu.

Schlüsselwörter: Psychopharmaka — Prüfungsmethodik — Psychophysiologische Reaktivität — Rating Scales — Stichprobenbildung.

Die Einführung der Psychopharmaka in die psychiatrische Therapie ist ein Fortschritt, dessen Bedeutung für das Los psychisch Kranker noch nicht endgültig beurteilt werden kann. Zu dieser Schlußfolgerung wird jeder kritische Beobachter kommen, wenn er die Lage psychiatrischer Patienten vor der Aera der Psychopharmakologie mit derjenigen unserer Tage vergleicht. Der gleiche Beobachter wird jedoch mit Erstaunen feststellen müssen, daß heute, 22 Jahre nach der Entdeckung der antipsychotischen Wirkungen von Chlorpromazin, trotz einer weltweiten Grundlagenforschung und trotz der Entwicklung zahlreicher Verfahren zur Erfassung der klinischen Wirkungen dieser Präparate, einfache Fragen der Therapie im Alltag ungelöst sind. Ich will nur eine herausgreifen:

Wenn ich einen schizophrenen Patienten vor mir habe, weiß ich zwar, daß er durch eine Behandlung mit Neuroleptika mehr Besserungschancen hat als mit andern therapeutischen Verfahren. Wenn ich aber begründen soll, weshalb ich ihm *ein bestimmtes* Neuroleptikum mit antipsychotischer Wirkung appliziere und nicht ein anderes, muß ich die Antwort schuldig bleiben.

Ähnlich, wenn nicht noch schlimmer, ist die Situation bei der Behandlung mit Antidepressiva, weil hier nicht nur das rechte Präparat, sondern auch noch die richtige Dosierung am einzelnen Patienten durch Versuch und Irrtum herausgefunden werden muß. Am bedenklichsten schließlich ist die Lage bei der Behandlung von sogenannten „Problempatienten“ mit Minortranquillizer, nicht nur wegen der bekannten hohen Placeboerfolgsrate, sondern wegen den verschwindend geringen klinischen Wirkungsunterschieden in dieser Präparatkategorie. Klinische Diagnose, Syndromdiagnose und Zielsymptome geben mir im Einzelfall nur sehr generelle Richtlinien für die Wahl einer bestimmten Präparatkategorie, jedoch keine präzisen Hinweise für das Präparat mit der größten Erfolgschance und die richtige Dosierung. Psychopathologische Kriterien haben nach den Studien von Raskin, Angst und Levine auch keine zuverlässige prognostische Bedeutung, d. h. sie lassen z. B. therapieresistente Fälle nicht voraussagen.

Diese Sachlage, welche die Kluft zwischen gesichertem Wissen und täglichem Handeln in der psychiatrischen Pharmakotherapie beleuchtet, steht nicht im Widerspruch zu der eingangs gemachten Feststellung. Es bleibt unbestritten, daß, statistisch gesehen, Schizophrene von Neuroleptika profitieren, daß Depressionen in einem hohen Prozentsatz durch Antidepressiva geheilt oder gebessert werden, und daß das Leiden bestimmter Patienten mit neurotischen und psychosomatischen Störungen durch Minortranquillizer vermindert wird. Die Schwierigkeiten einer vergleichenden Effizienzprüfung von Psychopharmaka beginnen eben nicht bei den klinischen Wirkungsunterschieden der großen Präparatkategorien, sondern bei den feineren Wirkungsdifferenzen *innerhalb* dieser Kategorien.

Soweit ich zu sehen vermag, sind die Hoffnungen, die man in weitgestreute, sorgfältig geplante, mit klinisch erprobten Skalen durchgeführte Doppelblindstudien gestellt hat, um endlich empirisch begründete Hinweise für eine zuverlässige Festlegung von feineren Wirkungsprofilen zu erreichen, bisher nicht erfüllt worden. Die meisten dieser Studien, die ein bestimmtes psychopathologisch definiertes Wirkungsprofil für ein Präparat festlegen, das über die Feststellung antipsychotisch/nichtantipsychotisch oder antidepressiv/nichtantidepressiv oder sedierend/nichtsedierend hinausgeht, werden durch andere Studien nicht bestätigt, oft gar widerlegt. Dies steht im krassen Gegensatz zu den pharmakologischen und biochemischen Untersuchungen am Tier, welche mit Zuverlässigkeit und Präzision feine Wirkungsunterschiede zwischen chemischen Strukturen *innerhalb* der klinisch definierten Präparatkategorien erkennen lassen.

Ich möchte im folgenden untersuchen, weshalb bisher die mit so großem Enthusiasmus und Arbeitsaufwand unternommenen Anstrengungen der Kliniker und Psychologen für die Entwicklung einer quantitativen, empirisch begründeten Psychopathologie zur Erfassung differentieller Wirkungsunterschiede von Psychopharmaka wenig erfolgreich waren, und einige Vorschläge zur Verbesserung der Methodik machen. Dazu möchte ich zwei Problemkreise berühren, die meines Erachtens wegen der Faszination der Forscher durch die große Zahl und durch die rechnerische Effizienz der Großrechner in letzter Zeit zu wenig beachtet wurden: Es sind dies

- I. Probleme der Merkmalserhebung pathologischen Verhaltens und
- II. Probleme der Stichproben, die in therapeutische Versuche eingehen.

I. Probleme der Merkmalserhebung pathologischen Verhaltens

Zweifellos stellen die heute allgemein eingeführten Rating-Scales und Symptomlisten für unsere Fragestellung einen Fortschritt gegenüber der früher üblichen, freien Deskription pathologischen Verhaltens und Erlebens dar. Im deutschen Sprachraum haben sich vor allem die Skalen von Lorr und von Overall sowie das AMP-Dokumentationssystem eingebürgert. Ohne auf die Vor- und Nachteile der einzelnen Beurteilungssysteme einzugehen, die von Mombour eingehend verglichen wurden, ist festzuhalten, daß sie alle prinzipiell einen Merkmalsraum definieren, in welchem sich pathologisches Verhalten und Erleben abbilden soll. Für die Abbildung des einzelnen Falles stehen einerseits Symptome oder Symptombereiche zur Verfügung, andererseits Stufen der Ausprägung. Im AMP-System sind die Symptome entsprechend dem klinischen Sprachgebrauch benannt und in einem Manual genauer umschrieben, in der Lorr-Skala sind sie als operationale Aussagen fixiert. Die Schwere eines Krankheitsbildes kommt in dem additiven Gesamtscore zum Ausdruck, den ein Patient erreicht. Die Besonderheiten eines Patientenkollektivs lassen sich als Mittelwerte eines Profils darstellen, das aus Symptombereichen zusammengestellt ist oder wie beim AMP-System als Mittelwerte in den Faktorenscores, der im Merkmalsraum faktorenanalytisch zusammengefaßten Syndrome (Baumann). Die Wirkung eines Präparates im Doppelblindvergleich mit einem andern Präparat wird durch die Beeinflussung einerseits der Gesamtscores, andererseits der einzelnen Symptombereiche bzw. der faktoriellen Syndrome bestimmt.

Wir besitzen zur Zeit keine besseren Verfahren, um überhaupt quantitative Aussagen über Psychopharmakawirkungen zu machen. Dabei sind methodische Mängel bei diesem Vorgehen augenfällig:

1. Mombour hat beispielsweise in einer vergleichenden Studie gezeigt, daß sowohl in der Lorr-Skala wie im AMP-System eine große Zahl von Symptomen angeführt ist, die in einer Stichprobe von 454 Patienten weniger häufig als in 10% der Fälle vorkommen. Synonyma in den beiden Skalen wurden ungleich häufig als vorhanden bewertet, z. B. im AMP-System seltener markiert als in der Lorr-Skala, obwohl die gleichen Patienten nach einem Interview mit den beiden Skalen zugleich dokumentiert wurden. Mombour führt dies auf die unterschiedliche Anordnung der Merkmale und auf die verschiedenen Instruktionen in beiden Systemen zurück.

2. Wir konnten zeigen, daß in Abhängigkeit von dem Krankheitsbild die Gesamtscores verschiedener Skalen vor und nach der Behandlung den Schweregrad nicht adäquat ausdrücken, weil die Skalen bei verschiedenen Schweregraden der Zustandsbilder Verschiedenes messen (Heimann u. Schmocker). Vor der Behandlung, also bei stärkerer Ausprägung der Krankheitssymptomatik, waren die Korrelationen der Gesamtscores zwischen verschiedenen Skalen der gleichen Patientenstichprobe niedriger als nach der Behandlung. Dies kann nur dadurch erklärt werden, daß diese sogenannten Meßverfahren bei verschiedenen Schweregraden der Erkrankung unterschiedliche Informationsanteile des gestörten Verhaltens und Erlebens berücksichtigen.

3. Diese Ergebnisse sind meines Erachtens ein Beweis dafür, daß wir durch das additive Vorgehen in der Symptomstatistik Symptome als Einheiten betrachten

und addieren, die in Wirklichkeit diese Bedeutung nicht haben können. Zwar sind grobe Unterschiede zwischen wirksamen und unwirksamen Präparaten auf diese Weise durchaus feststellbar. Der ganze Aufwand einer massenstatistischen Erfassung einzelner Merkmale wird jedoch gerade betrieben, um feinere Unterschiede quantitativ und zuverlässig definieren zu können.

4. *Häufigkeit* und *Spezifität* eines Merkmals stehen in einem umgekehrten Verhältnis. Das zeigten die Merkmalsverteilungen, welche Mombour publizierte. Die folgende Tabelle 1, welche eine Auswahl aus den AMP-Merkmalen der Arbeit von Mombour gibt, zeigt in der ersten Gruppe Symptome, die relativ spezifisch sind, aber selten vorkommen, beispielsweise „Gedankenentzug“, „Störungen der Ich-Identität“. Sie sind für schizophrene Erleben sehr typisch, aber, wie die vorliegende Statistik beweist, doch relativ selten. In der zweiten Gruppe sind Merkmale zusammengefaßt, die häufig vorkommen, aber dafür sehr unspezifisch sind, z. B. das Merkmal „innerlich unruhig“ oder „mißtrauisch/feindselig“. In der dritten Gruppe sind schließlich Merkmale angeführt, die ein ganz unspezifisches psychisches Minussyndrom ergeben, eine Herabsetzung des vitalen Wohlbefin-

Tabelle 1. Häufigkeit der AMP-Items bei verschiedenen Diagnosen (in Prozent, aufgerundet^a auf ganze Zahlen)

Item- Nr.	Psychischer Befund	Organ. Psych.	Schizophr. u. ä.	Depres- sionen	Manie etc.	Neurosen	Organ. Wesens- änderun- gen
		(n = 36)	(n = 152)	(n = 125)	(n = 6)	(n = 118)	(n = 15)
37	Wahndynamik	3	26	1	0	0	0
38	Systematisierter Wahn	6	19	2	0	0	7
48	Stimmenhören	17	22	1	0	3	0
54	Störungen der Ich- Identität	3	15	1	0	2	0
56	Autismus	8	13	2	0	3	7
58	Gedankenentzug	0	6	0	0	0	0
60	sonstige Fremdbein- flussungsergebnisse	6	18	1	0	0	0
75	Läppisch	3	13	2	0	7	13
66	Deprimiert/Traurig	36	34	82	17	48	60
68	Ängstlich	25	29	57	0	32	27
71	Mißtrauisch/Feindselig	19	39	14	33	27	33
73	Innerlich unruhig	33	37	58	50	48	20
87	Antriebsam	31	44	39	0	25	33
108	Ablehnung der Behandlung	17	24	10	33	10	20
103	Kontakt vermindert	28	59	44	0	42	47
105	Krankheitsgefühl	42	44	58	0	43	40
115	Sexualität vermindert	11	24	38	17	25	20
124	Einschlafstörungen	47	31	62	0	44	27
131	Appetit vermindert	36	28	50	33	28	13
150	„Kopfdruck“ o. ä.	36	20	23	17	28	40

^a Nach W. Mombour: Arch. Psychiat. Nervenkr. 219, 133–152 (1975).

dens. Sie sind wiederum häufig und verteilen sich so wie diejenigen der zweite Gruppe. Wenn man diese empirischen Verteilungen der Merkmale in Betracht zieht, wird man statistischen Analysen, welche die Merkmale als gleichwertige Einheiten behandeln, mit einiger Skepsis begegnen. Die gleichen Verteilungsunterschiede zwischen Diagnosen ergeben sich nach Mombour auch für die Lorr-Skala.

Man könnte nun einwenden, daß ein besseres Beobachtungstraining diese offensichtlichen Mängel beheben könnte, nämlich dadurch, daß Beobachter durch ein systematisches Training präzisere Vorstellungen entwickeln, wann ein Phänomen mit einem bestimmten Sprachsymbol belegt wird und wann nicht. Obwohl ein solches Training für die Entwicklung einer gemeinsamen Beobachtungssprache von unschätzbarem Wert ist, darf man nicht übersehen, daß die hier aufgezeigten Unterschiede zwischen Merkmalen *unabhängig vom Beobachterparameter* sind. Sie können nicht behoben werden, weil sie durch einen *zweiten Parameter* jedes in Frage stehenden Merkmals festgelegt sind, nämlich durch *seine Häufigkeit in einer Stichprobe verschiedener klinischer Zustandsbilder*.

Wenn man unseren Überlegungen bis hierher gefolgt ist, wird man zur Einsicht gelangen, daß der *Vorteil* der gebräuchlichen Dokumentationssysteme, nämlich einen weiten Bereich klinischer Psychopathologie zu erfassen, mit dem *Nachteil* erkauft wird, daß Merkmale mit nichtvergleichbarer deskriptiver Bedeutung bzw. diagnostischem Stellenwert als statistische Einheiten behandelt werden. Will man diesen Nachteil vermeiden, gibt es keine andere Möglichkeit, als aus der Fülle gebräuchlicher, deskriptiver Merkmale der Psychopathologie jene auszulesen, deren empirische Verteilungscharakteristika für eine bestimmte Patientenkategorie zeigen, daß sie sich für die Beurteilung der Schwere eines psychopathologischen Zustandsbildes eignen.

Eine solche Merkmalselektion ist auf der Basis des *probabilistischen Testmodells nach Rasch* möglich. Das Modell geht von einer Matrix aus, welche in den Spalten die Patienten nach der Häufigkeit vorhandener Merkmale einer Symptomenliste ordnet, in den Zeilen die Merkmale. Nach diesem Modell ist die Schwere eines Zustandsbildes operational definiert durch die Häufigkeit vorhandener Symptome und die Merkmalsverteilung als Funktion der Krankheitsschwere überblickbar. Merkmale, die zur statistischen Deskription geeignet sind, weisen eine logarithmische Funktion in bezug auf die Krankheitsschwere auf. Das Modell kann dreidimensional erweitert werden, wenn in der dritten Dimension der Ausprägungsgrad der Merkmale aufgetragen wird (vgl. dazu G. Fischer).

Ich möchte dies an einem praktischen Beispiel erläutern, einem Feldversuch mit 6 Internisten, 11 Allgemeinpraktikern und 5 Psychiatern, den wir mit Bethge, Boller und Stoll in der Umgebung von Darmstadt durchgeführt haben. Es handelt sich um eine Doppelblindstudie mit Mepiprazol in der Indikation eines Tranquillizers gegen Placebo. Die Patienten wurden unter anderem mit Symptomenliste von elf Symptomen beurteilt, die zur Erfassung der zu behandelnden Störungen geeignet erschienen. In einer 5stündigen, gemeinsamen Vorbesprechung wurde die Indikationsstellung und die Bedeutung der Symptome mit den Kollegen eingehend diskutiert, welche alle vorher nie an einer ähnlichen Studie teilgenommen hatten.

Tabelle 2. Elf Symptome für Feldversuch mit Tranquiliizer: 1. Kolonne: + = Symptome, die dem Modell von Rasch entsprechen. 3. Kolonne: Faktorenstruktur aller Ärzte zusammen. 4. Kolonne: × = Symptome, die in den Faktorenstrukturen der drei Arztkategorien überzufällige Unterschiede zeigten (Analyse nach Fischer-Roppert)

	Symptome	Varimaxfaktoren		
		I	II	III
+	Gereiztheit	74	-05	03
	Nervöse Erschöpfung	64	19	-06
	Motorische Unruhe	66	16	-09
	Erregtheit	80	03	10
	Innere Unruhe	70	10	20
	Deprimiertheit	25	56	21
+	Somatische Symptome	02	09	94
+	Ängstlichkeit	08	58	26
+	Selbstunsicherheit	09	87	-05
+	Kontaktstörungen	12	74	13
+	Schlafstörungen	33	23	01

Eine Analyse der Merkmalsverteilung nach Abschluß der Studie ergab, daß von den elf Symptomen nur sechs den Anforderungen der Homogenität und den Modellvoraussetzungen nach Rasch entsprachen. Sie sind in der vorstehenden Tabelle 2 mit einem Kreuz in der ersten Kolonne angezeigt. Was ist durch eine Beschränkung auf diese sechs Symptome gewonnen? Krankheitsschwere und Charakteristika der Stichprobe können mit ihnen adäquat beschrieben werden, ebenso die Besserung durch die Behandlung. Das Ziel ist gewissermaßen weniger weit gesteckt, aber zuverlässiger erreichbar. Denn umfangreichere Skalen, wie die eingangs erwähnten, machen prinzipiell die gleichen Voraussetzungen, ohne sie, wie wir gesehen haben, mit den gebräuchlichen Gesamt- oder Faktorenscores erfüllen zu können. Dennoch ist es meines Erachtens durchaus sinnvoll, die Patienten einer Stichprobe mit umfangreicheren Skalen zu dokumentieren. Die zur Erfassung psychopharmakologischer Wirkung geeigneten Merkmale können dann sekundär nach der Methode von Rasch aus den empirischen Verteilungen herausgesucht werden. Mit ihnen kann hierauf die therapeutische Wirkung zuverlässiger beschrieben werden als nach dem herkömmlichen Vorgehen.

Die Resultate unseres Feldversuches haben jedoch noch ein anderes Problem enthüllt, das ich nur kurz streifen möchte, nämlich die *semantische Unschärfe* der deskriptiven Merkmale in der Psychopathologie. Wenn man die Summenscores der Merkmalslisten für alle Untersucher und alle 384 Patienten der Doppelblindstudie zwischen der Placebogruppe und der Medikamentgruppe vergleicht, ergibt sich statistisch kein Unterschied zwischen den Behandlungen; ein bemerkenswertes Ergebnis, das antipsychiatrische Hitzköpfe zu voreiligen Interpretationen verleiten könnte. Die erste Abbildung zeigt jedoch die Resultate der kovarianzanalytisch angepaßten Mittelwerte für die Summenscores der Symptomliste, nach den drei Ärztegruppen gesondert aufgetragen. Hier wird deutlich, daß die Internisten und die Allgemeinpraktiker keine signifikanten Unterschiede beobachteten, wohl aber die Nervenärzte. Nur bei den beiden erst-

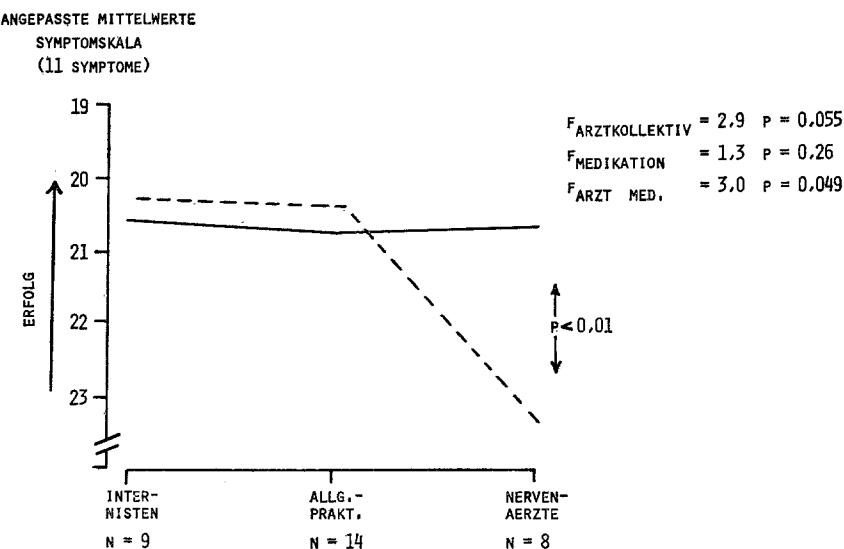


Abb. 1. Kovarianzanalytisch angepaßte Mittelwerte für die Symptomsummenwerte (elf Symptome nach Behandlung) bei drei Arztgruppen ($N = 384$ Patienten mit vollständigen Angaben)

genannten Gruppen war die Placeborate so außerordentlich hoch, was somit *nicht einer realen Placebowirkung entspricht*, sondern *durch die Unschärfe der Beobachtung bedingt ist*. Durch unsere Versuchsanordnung konnte hier der für die Merkmale ebenso wesentliche *Beobachterparameter* demonstriert werden.

Daß tatsächlich die drei Beobachtergruppen unter dem gleichen Sprachsymbol Verschiedenes verstanden, geht schließlich aus Tabelle 2 hervor.

In der letzten Kolonne sind durch ein \times die Symptome angegeben, die in den Faktorenstrukturen der drei Ärztekategorien überzufällige Unterschiede zeigten. Besonders bemerkenswert ist dies z. B. für die Symptome „Deprimiertheit“, „Ängstlichkeit“ und „Schlafstörungen“. Es ist selbstverständlich, daß solche Unterschiede des semantischen Verständnisses durch ein besseres Training der Beobachter mit Hilfe von Videotape ausgeglichen werden können.

Die zweite Abbildung zeigt, daß ein Lernprozeß bei den Internisten und Allgemeinpraktikern während des Versuches schon spontan in Gang kam. Es ist daran zu erkennen, daß die Placeboerfolgsrate absinkt, und die Mepiprazolerfolgsrate ansteigt, wenn man die erste Hälfte der behandelten Patienten mit der zweiten Hälfte vergleicht.

Diese Probleme der Merkmalserfassung pathologischen Verhaltens und Erlebens sind die vernachlässigte Grundlage aller statistischen Aussagen über die Wirkung von Psychopharmaka im klinischen Bereich. Ihre Berücksichtigung und erfolgreiche Bewältigung erfordert eine Denkarbeit, die uns der Computer nicht abnehmen kann.

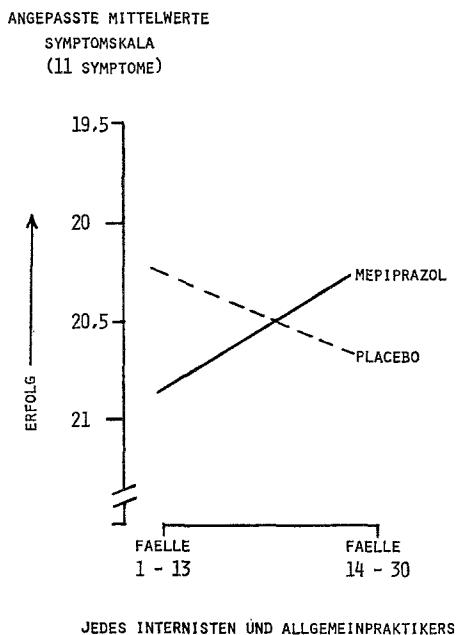


Abb. 2. Unterschiede in der Bewertung von Placebo- und Medikamentenerfolg während des Versuchs

II. Probleme der Stichproben, die in therapeutische Versuche eingehen

Der zweite Problemkreis, nämlich die Stichprobenbildung für therapeutische Versuche, soll nur kurz berührt werden. Es ist hier auf den Informationsverlust hinzuweisen, den wir durch die Bildung von Mittelwerten in Kauf nehmen müssen:

1. Nach Epstein zeigen beispielsweise Schizophrene in psychophysiologischen Untersuchungen verglichen mit Normalen entweder zu starke oder zu schwache Reaktionen auf äußere Stimuli, oder einzelne Patienten reagieren einmal zu stark, ein andermal zu schwach. Wenn man Mittelwerte von einer Stichprobe von Schizophrenen mit Normalen vergleicht, können diese entweder zu hoch oder zu tief liegen oder sich von denjenigen Normalen nicht unterscheiden, je nach dem Verhältnis der Typen, die mehr oder weniger zufällig in die Stichprobe eingehen. Dieselben Konsequenzen, nämlich eine *Inhomogenität vom psychophysiologischen Standpunkt* ergeben sich, wenn nach klinischen Gesichtspunkten Patienten in Stichproben aufgenommen werden, mit welchen man ein neues Neuroleptikum gegen ein Standardpräparat testen will. Hinter der klinischen Diagnose „Schizophrenie“ können sich psychophysiologisch unterschiedliche Typen verbergen, was für die Inkonsistent feinerer Wirkungsprofile von Psychopharmaka verantwortlich sein kann.

2. In der Depressionsforschung beginnen wir erst die Bedeutung der circadianen Rhythmen biologischer Funktionen zu berücksichtigen. Wenn die wesentliche Störung in einer Desynchronisierung dieser Rhythmen liegt (Pflug), brauchen

sich Mittelwerte von Stichproben Depressiver überhaupt nicht von jenen Normaler zu unterscheiden, denn das Krankhafte zeigt sich in der Phasenverschiebung, die durch die Mitteilung verlorengeht.

3. Psychophysiologische Untersuchungen Depressiver zeigen jedoch zwischen Patienten noch viel krassere Unterschiede, welche durch die klinische Merkmalserhebung überhaupt nicht erfaßt werden. Das sei am Beispiel der psychophysiologischen Resultate zweier depressiver Patientinnen abschließend demonstriert (vgl. dazu auch H. Heimann). Beide Patientinnen sind etwas über 30 Jahre alt, beide litten klinisch an einem depressiven Zustandsbild mit den klassischen Symptomen. Beide wiesen in der Anamnese frühere depressive Phasen auf. Sie wären aufgrund ihrer klinischen Daten zusammen in eine Stichprobe zur Prüfung eines Antidepressivums eingegangen.

Abb.3 zeigt die Ergebnisse beider Patientinnen im Konditionierungsexperiment. Während Patientin A auf das Geräusch (unkonditionierter Reiz) in allen Variablen deutlich reagiert, bleibt Patientin B mit Ausnahme der Alphablockierung reaktionslos.

Abb.4 zeigt die Situation im Moment, wo das Geräusch nicht wie erwartet gegeben wird, also die Situation der Konditionierung. Patientin A zeigt eine rasche Konditionierbarkeit, Patientin B überhaupt keine Reaktion.

In Abb.5 ist die Reaktion auf einen vergleichbaren verbalen Stimulus dargestellt. Patientin A reagiert in allen Variablen deutlich auf das Reizwort „Schmerz“, während Patientin B auf das Reizwort „Kummer“ keine Reaktionen zeigt.

Die psychophysiologische Reaktivität dieser beiden Patientinnen ist somit völlig unterschiedlich, obwohl sie klinisch ein depressives Zustandsbild zeigen. Es wurde von erfahrenen Kollegen im Falle der Patientin A als endogen, im Falle der Patientin B als vorwiegend neurotisch beurteilt.

Die Resultate der psychophysiologischen Untersuchungen und die Thearbeitserfolge bewogen uns, die Klassierung der beiden Patientinnen gerade umzukehren. Patientin A hat auf zwei verschiedene Antidepressiva so paradox reagiert, daß die Behandlung abgebrochen werden mußte. Patientin B erhielt zunächst ungenügende Dosen verschiedener Antidepressiva, sprach aber auf eine genügend hoch dosierte Kombinationsbehandlung von Dibenzepin und Nortriptylin rasch mit einer völligen Remission der depressiven Symptomatik an.

Beide Patientinnen hätten in einer Stichprobe Depressiver zu Mittelwertbildungen in Verhaltensskalen beigetragen, Mittelwerten, in welchen diese riesigen Unterschiede der psychophysiologischen Reaktivität überhaupt nicht erfaßt werden.

Solche Untersuchungen zeigen, daß wir in der klinischen Psychopharmakologie erst am Anfang stehen. Die bisherige Forschungsstrategie großer Stichproben welche auf Grund von Verhaltensmerkmalen zusammengestellt werden, kann die praktisch relevante Frage gar nicht beantworten, die wir eingangs stellten, nämlich: Welches ist das richtige Präparat für den richtigen Patienten?

Wir müssen wieder zum einzelnen Patienten zurück und ihn auf verschiedenen Ebenen: Verhaltensmerkmale, Erlebnisbereich, psychophysiologische Reaktivität etc., untersuchen, wobei der zeitliche Verlauf einer besonderen Berücksichtigung bedarf. Dazu sind neue statistische Modelle für die Einzelfallforschung und Zeit-

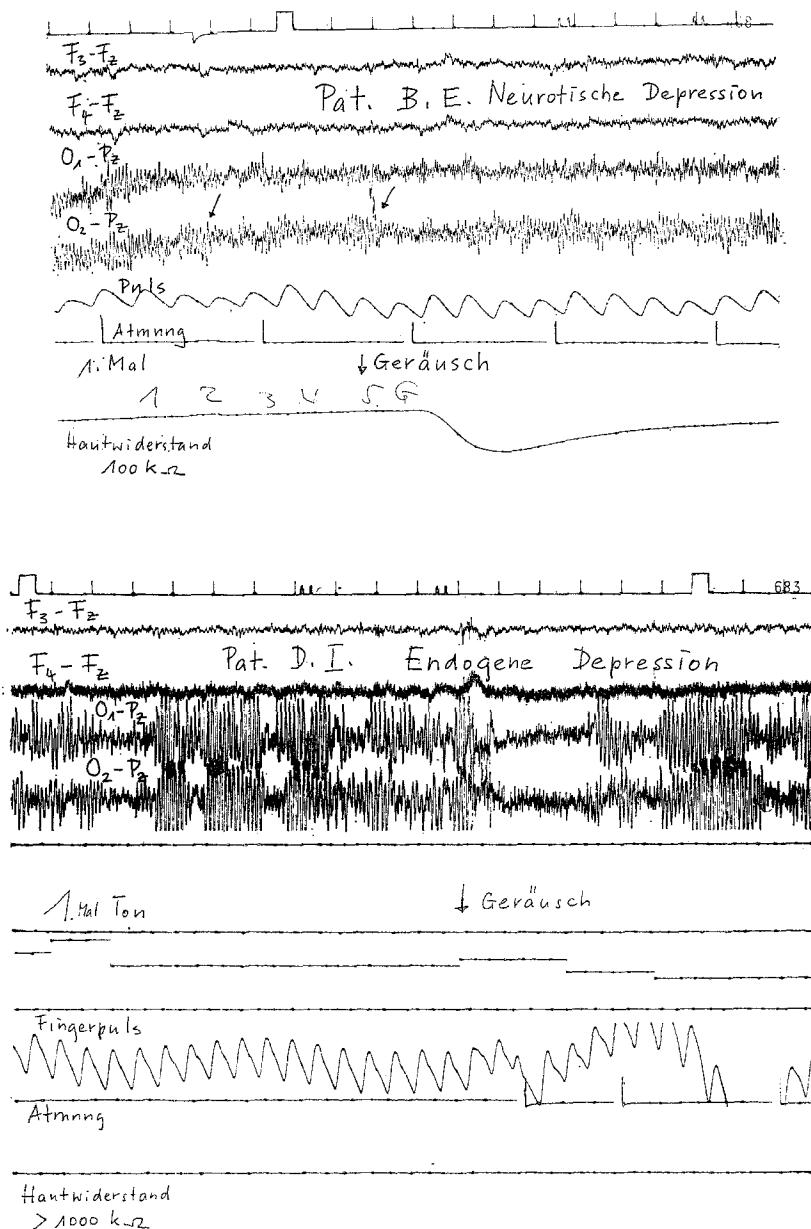


Abb.3. Resultate der polygraphischen Untersuchung von zwei depressiven Patientinnen im Konditionierungsexperiment, Kanal 1–4: 2 frontale und 2 parieto-occipitale Ableitungen des Elektrencephalogramms. Kanal 5–8: Fingerpuls, Atmung, Hautwiderstand. Reaktion auf erstmalige Darbietung eines starken Geräusches (96 dB) nach ankündigendem Ton bzw. Zählen. Weiteres: siehe Text

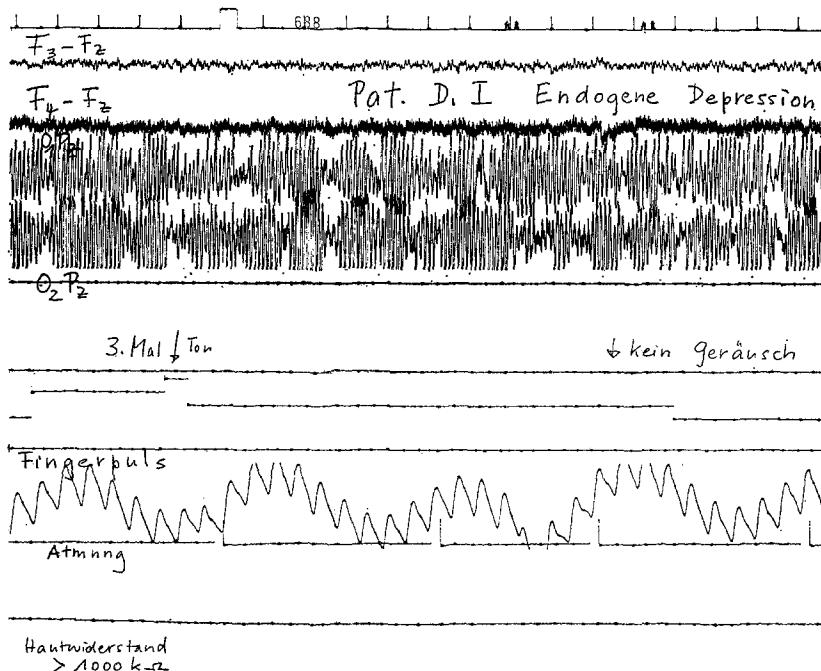
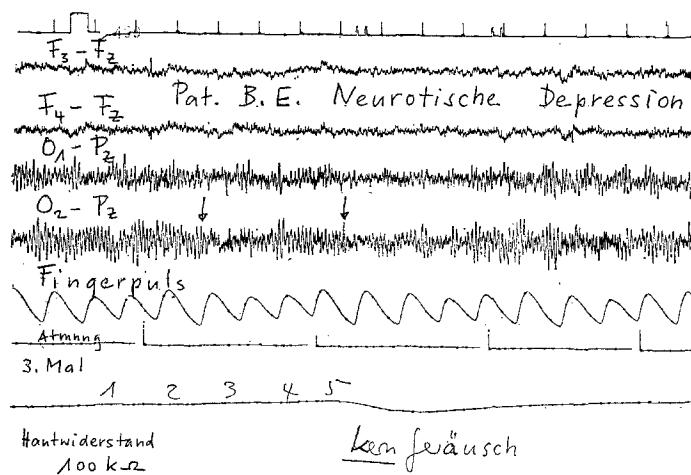


Abb.4. Resultate der polygraphischen Untersuchung von zwei depressiven Patientinnen im Konditionierungsexperiment. Kanal 1–4: 2 frontale und 2 parieto-occipitale Ableitungen des Elektroenzephalogramms. Kanal 5–8: Fingerpuls, Atmung, Hautwiderstand. Reaktion auf ausbleibendes Geräusch nach 3. ankündigendem Ton oder Zählen. Weiteres: siehe Text

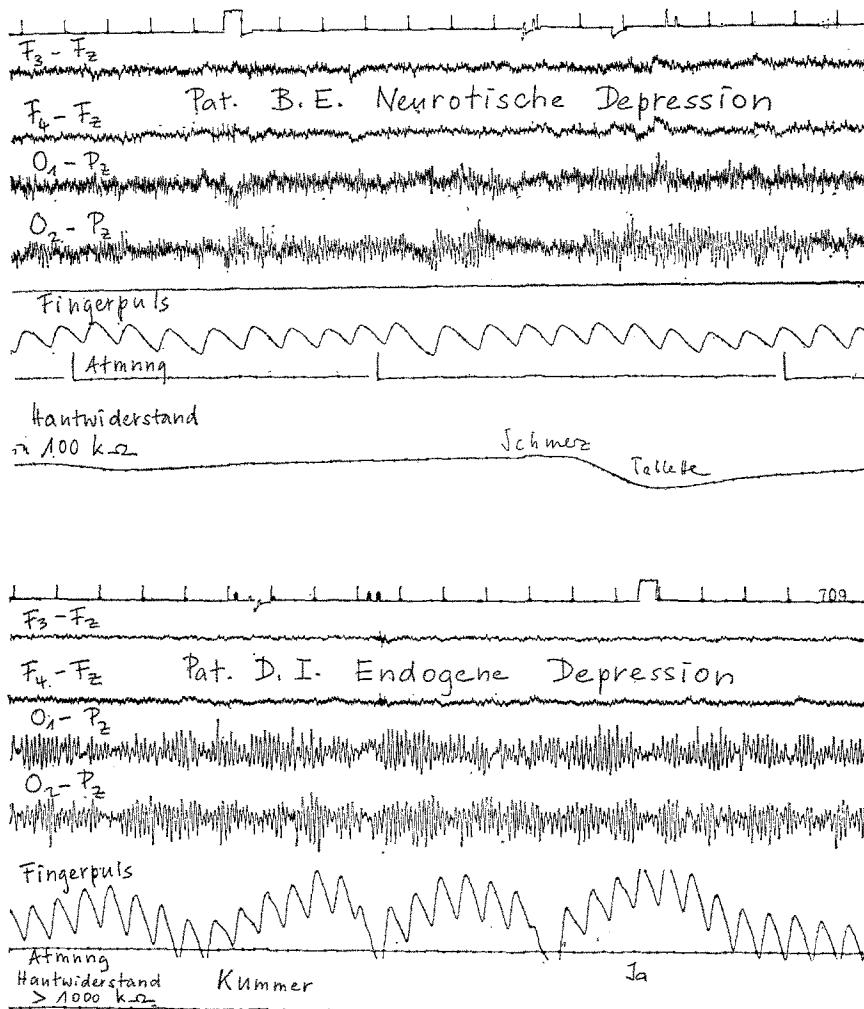


Abb. 5. Resultate der polygraphischen Untersuchung von zwei depressiven Patientinnen während des Jüngschen Assoziationsversuches. Kanal 1–4: 2 frontale und 2 parieto-occipitale Ableitungen des Elektrencephalogramms. Kanal 5–8: Fingerpuls, Atmung, Hautwiderstand. Reaktion auf Darbietung eines emotionalen Wortes (Kummer bzw. Schmerz). Weiteres: siehe Text

reihenanalyse anzuwenden, welche die Psychopharmakologie wieder näher an die biographisch orientierte psychodynamische Betrachtungsweise in der Psychiatrie heranführt.

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.: Das Dokumentations-System der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMP). Arzneimittel-Forsch. 19, 399–405 (1969)

Angst, J., Baumann, U., Hippius, H., Rothweiler, R.: Clinical aspects of resistance to Imipramin therapy. *Pharmakopsychiat.* **7**, 211—216 (1974)

Baumann, U., Rothweiler, R., Scheidegger, P.: Methodische Probleme bei Psychopharmakostudien mit dem AMP-System. Vortrag auf dem Symposium der AGNP in Nürnberg, Oktober 1973

Epstein, S., Colemann, M.: Drive theories of schizophrenia. *Psychosom. Med.* **32**, 113—140 (1970)

Fischer, G.: Einführungen in die Theorie psychologischer Tests. Bern, Stuttgart, Wien: H. Huber 1974

Fricke, R.: Zeittheorie des Rasch-Modells und einige Erfahrungen bei der Anwendung dieses Modells auf Schulleistungs-Tests. *Z. erziehungswissenschaftl. Forsch.* **6**, 3—23 (1972)

Heimann, H.: Therapy-Resistant depressions: Symptoms and syndromes. *Pharmakopsychiat.* **7**, 139—144 (1974)

Heimann, H., Boller, W., Stoll, K. D., Junebloed, F.: Drug-placebo-difference in neurotic outpatients of internists, general practitioners and psychiatrists using a symptom checklist analysed by methods based on the probabilistic test model of Rasch. CINP-Congress Paris 1974 (im Druck)

Heimann, H., Schmocker, A.: Zur Problematik der Beurteilung des Schweregrades psychiatischer Zustandsbilder. *Arzneimittel-Forsch. (Drug Res.)* **24** 1004—1006 (1974)

Lorr, M.: Inpatient multidimensional psychiatric scale. Consulting Psychologists Press, Palo Alto, Calif. 1963

Mombour, W.: Symptomhäufigkeiten bei psychiatrischen Erkrankungen. *Arch. Psychiat. Nervenkr.* **219**, 133—152 (1974)

Overall, J. E., Gorham, D. R.: The brief psychiatric rating scale. *Psychol. Rep.* **10**, 799—812 (1962)

Pflug, B.: Depression und Schlafentzug. Neue therapeutische und theoretische Aspekte. *Habilitationsschrift, Tübingen* 1973 (im Druck)

Raskin, A.: The prediction of antidepressant drug effects: Review and critique. *Psychopharmacology, a review of progress*. Ed.: D. H. Efron. Public Health Service Publication No. 1836, 757 (1968)

Prof. Dr. H. Heimann
Psychiatrische Universitätsklinik
D-7400 Tübingen
Bundesrepublik Deutschland