

# DREI FORDERUNGEN ZUR PRÜFUNG DER HOMOGENITÄT VON ZENTREN BEI MULTICENTER-STUDIEN

R. KOHNEN, H.-P. KRÜGER, G. A. LIENERT  
Universität Erlangen-Nürnberg

## Abstract

Es wird argumentiert, daß bei der Prüfung der Homogenität von Zentren in mehrfaktoriellen Multicenter-Studien vor allem entscheidend ist, daß sich die Zentren in den faktoriellen Variablen nicht unterscheiden, deren Effekte in der späteren Auswertung interpretiert werden sollen. Dazu wird folgendes Procedere vorgeschlagen: (1) überprüfe, ob in allen Zentren die Randomisierungsbedingungen eingehalten wurden, (2) überprüfe, ob die Zentren homogen bezogen auf die eingebrachten Schichtvariablen (z.B. Geschlecht, Indikation) sind und (3) überprüfe, ob die Zentren homogen sind bezogen auf die Wechselwirkungen zwischen den eingebrachten faktoriellen Variablen. Es wird gezeigt, daß vor allem Forderung (3) für die Interpretationsfähigkeit der Ergebnisse von entscheidender Bedeutung ist. Möglichkeiten der statistischen Prüfung dieser Voraussetzungen werden an einem Beispiel aus der klinischen Prüfung eines Psychopharmakons vorgestellt.

## 1. Das Forschungsinstrument "Multicenter-Studie"

Multizentrische Studien (MZS) sind Untersuchungen, die nach ein und demselben Versuchsplan an verschiedenen Orten simultan oder sukzessiv durchgeführt worden sind. Voraussetzung für die Zusammenfassung der einzelnen Zentren und damit für die Interpretierbarkeit der Studie als ganzer (externe Validität sensu CAMPBELL & STANLEY, 1963) ist die Homogenität der Zentren. Eine wichtige Technik zur Sicherung der Qualität von MZS liegt in der ex post-evaluation der Homogenität. Dabei wird an den vorliegenden Daten untersucht, inwieweit statistisch zu rechtfertigen ist, die Hypothese der Homogenität der Zentren beizubehalten.

Die erste wichtige Unterscheidung bei der Homogenitätsprüfung liegt darin, ob die Studie faktoriell oder afaktoriell angelegt ist (zu der Nomenklatur siehe LIENERT & KRAUTH, 1979). Bei den faktoriellen Studien wird mindestens ein Merkmal (in der Regel Behandlungen) systematisch variiert. Dazu werden die Ptn eines jeden von mehreren Zentren nach Los (oder nach einem Randomisierungsplan) den verschiedenen Behandlungen zugeordnet. Werden die Ptn vor der Behandlungsbeurteilung nach bestimmten klinisch relevanten Merkmalen (wie Geschlecht oder Diagnose) geschichtet, so resultieren Untersuchungspläne, die neben einem randomisierten Behandlungsfaktor noch einen oder mehrere Schichtungsfaktoren implizieren.

## 2. Forderungen an die Homogenität der Zentren

Folgt man CAMPBELL & STANLEY (1963), ist die Vorbedingung für die Interpretation eines Ergebnisses die interne Validität einer Untersuchung, die dann gegeben ist, wenn für den Effekt nur die systematisch variierte Behandlung verantwortlich gemacht werden kann. Die wichtigste experimentelle Technik dafür ist die Randomisation der Behandlungen, die gewährleistet, daß kein anderes Merkmal als die variierte Behandlung auf die abhängige oder Zielvariable (Observable) einwirken kann. Damit ergibt sich

**F o r d e r u n g 1:** die Behandlung muß in jedem Zentrum den Ptn nach Zufallsbedingungen zugeteilt (alloziert) worden sein (RANDOMISATIONS - Kriterium).

In der Praxis von Medikamentprüfungen wird diesem Desiderat (1) durch das Aufstellen eines Randomisierungsplans, (2) die doppelblinde Verabreichung und (3) durch eine sichere Verschlüsselung Sorge getragen. Zentren, die diesen 3 Randomisationsbedingungen nicht genügen, müssen aus der Auswertung eliminiert werden.

Methodisch Neues gegenüber der Forderung (1) bringt der Fall, daß Ptn-Merkmale (organismic factors wie Geschlecht, Diagnosen oder Ausgangswerte) als eigene Schichtungsvariablen eingebracht werden. Wird das getan, ergibt sich

**F o r d e r u n g 2:** es muß angenommen werden dürfen, daß die Zentren homogen in bezug auf die Schichtvariablen sind (STRATIFIKATIONS - Kriterium).

Beide Forderungen (1) und (2) resultieren aus der gleichen Überlegung: Wenn eine Fragestellung faktoriell untersucht wird, dürfen keine Zentrumsmerkmale auf die zu untersuchenden Merkmale einwirken können. Nur unter dieser Bedingung kann sichergestellt werden, daß Effekte auf den Faktoren (an denen der Untersucher interessiert ist), nur auf diese Faktoren zurückgehen und nicht Eigenheiten der Zentren darstellen.

Soll letzteres ausgeschlossen werden, ergibt sich sofort eine neue Forderung, die bislang noch nicht aufgestellt wurde. Werden mehr als zwei Behandlungen (Faktoren) eingeführt, resultieren neben den Haupteffekten auch noch die Wechselwirkungen der Faktoren. In diesen Wechselwirkungen drückt sich im Fall von Schichtungs-faktoren die vom Untersucher hypostasierte differentielle Wirksamkeit seiner Behandlung aus. Damit ist die Wechselwirkung von eigener klinischer Aussagekraft. Für die interne Validität der Untersuchung muß dann aber auch klargestellt sein, daß die Zentren bezogen auf die Wechselwirkungen homogen sind. Das ergibt die

**F o r d e r u n g 3:** Bei mehrfaktoriellen Untersuchungsplänen müssen die Zentren homogen in Bezug auf die Wechselwirkungen der Faktoren sein (INTERAKTIONS - Kriterium).

### 3. MZS-Auswertung am Beispiel der klinischen Prüfung eines Psychopharmakons

Bei den Daten handelt es sich um eine multizentrischen Doppelblindstudie, in der Wirksamkeit und Verträglichkeit einer einmaligen morgendlichen Gabe von 20 mg Prazepam (Demetrin<sup>R</sup>) im Vergleich zu 10 mg Diazepam über 14 Tage verglichen wurde. Beteiligt waren 4 niedergelassene Psychiater (Zentren). Entsprechend Teil 2 werden in die Homogenitätsprüfung nur die Variablen einbezogen, deren Effekte in der Auswertung bestimmt werden. Als wirkungsbestimmender Faktor geht (1) die Medikation (randomisierter treatment Faktor) in die Untersuchung ein. Als wirkungsmodifizierende Faktoren wurden aus klinischen Gründen betrachtet (2) das Geschlecht und (3) die Indikation als Schichtungsfaktoren.

#### 3.1. Das Randomisationskriterium

Aus unserem Forderungskatalog ergibt sich damit der Ablauf der Homogenitätsprüfung in drei Schritten:

1. Überprüfe, ob in jedem Zentrum die Randomisationsbedingungen eingehalten wurden. Diese Prüfung ist stichprobenunabhängig zu vollziehen.

Das Randomisierungskriterium der Forderung 1 darf als erfüllt gelten, wenn in einem Untersuchungsplan mit gleichen Anteilen von je  $n(Z)/2$  Ptn für Verum- und Placebogruppe die  $Z$  Zentren etwa je zur Hälfte V- und P-Ptn enthalten. Die Wahrscheinlichkeit der Einhaltung der Randomisierungsbedingung ergibt sich im vorliegenden Fall zweier Medikamente mit gleichen Applikationshäufigkeiten aus der Binomialverteilung. Tabelle 1 gibt die Medikamentverteilung in den vier Zentren wieder.

	DIA	PRA	I	SUM
Z1	4	3	I	7
Z2	5	5	I	10
Z3	16	15	I	31
Z4	5	5	I	10
SUM	30	28	I	58

Tabelle 1: Die Verteilung des randomisierten Behandlungsfaktors Medikation mit den Stufen Diazepam (DIA) und Pranzepam (PRA) auf die vier Zentren Z1 - Z4.

Da Pranzepam und Diazepam gleich häufig in der Untersuchung waren, sind ihre Auftretenswahrscheinlichkeiten  $p(P) = p(D) = 1/2$ . Die Überschreitungswahrscheinlichkeit ergibt sich damit aus der Binomialverteilung mit den Parametern  $p$  und  $N$  zu

$$P = 1 - \sum_{x=0}^n \binom{n}{x} \cdot p(M_1)^x \cdot p(M_2)^{n-x}$$

wobei  $n$  = Ptn-Zahl des Zentrums,  $x$  = Häufigkeit des häufiger gegebenen Medikaments,  $p = p(M_1)$  = Häufigkeit des Medikaments 1,  $q = 1-p = p(M_2)$  = Häufigkeit des Medikaments 2. Bei großem  $N$  kann approximativ nach der Normalverteilung geprüft werden.

Entscheidend bei diesem Test ist, daß *n i c h t* über einen  $Z_{xm}$  - Felder Chi-Quadrat-Test ( $Z$  Zentren,  $m$  Medikamente) geprüft wird. Gleichgerichtete Abweichungen von den Randomisationsbedingungen könnten sonst unerkannt bleiben. Der Chi-Quadrat-Test prüft lediglich auf Gleichverteilung der Zentren, indem er die Randsummen als Erwartungswertschätzer benutzt. Hier liegen die Erwartungswerte aber bereits a priori fest.

### 3.2. Das Stratifikationskriterium

Im zweiten Schritt der Auswertung erfolgt die Prüfung der Schichtungsfaktoren Geschlecht und Indikation:

2. Überprüfe, ob die Zentren hinsichtlich der Verteilung der Schichtungsfaktoren homogen sind.

Nimmt man unter  $H_0$  der Schichtungshomogenität der Zentren von  $N$  Ptn aus  $Z$  Zentren an, daß die  $Z$  Zentren Zufallsstichproben aus den Subpopulationen der Schichtungsklassen von Ptn sind, dann ist der geeignete Test zur asymptotischen Beurteilung der Homogenität PEARSON's  $Z \times D$  - Felder -  $X^2$  - Test mit

$$X^2 = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \text{ mit } df=(Z-1)(S-1), \text{ wobei } e_{ij} = n_{i.} \cdot n_{.j} / N$$

die unter  $H_0$  erwarteten Frequenzen und  $f$  die beobachteten Frequenzen in der aus dem Zentrum  $i$  und der Schichtklasse  $j$  gebildeten Feld der entsprechenden Kontingenztafel bezeichnen. Ist PEARSON's  $X^2$  nach (3) auf einer vorgegebenen Stufe signifikant, dann sind die Zentren in Bezug auf die  $S$  Schichtklassen heterogen. Zur Prüfung, welches Zentrum eliminiert werden muß, um das Homogenitätskriterium ex post zu erfüllen, schiebe man  $Z$  simultane  $2 \times S$  - Felder -  $X^2$  - Tests nach, in welchen jeweils ein Zentrum  $j$  den gepoolten  $Z-1$  Zentren gegenübergestellt wird und eliminiere jene Zentren, die ein signifikantes  $X^2$  ergeben. Als Beispiel wählen wir die Verteilung des Merkmals Diagnose in den 4 Zentren der Studie (Tabelle 2).

	NEU	SIT	PER	MIS	SUM
Z1	4	1	2	0	I 7
Z2	6	4	0	0	I 10
Z3	14	14	3	0	I 31
Z4	1	3	0	6	I 10
SUM	25	22	5	6	I 58

Tabelle 2: Die Verteilung der vier Diagnosen "neurotische Störung" (NEU), "situationsbedingte Störung" (SIT), "persönlichkeitsbedingte Störung" (PER) "Mischdiagnose" (MIS) auf die vier Zentren.

Das  $X^2 = 38,99$  bei  $df=9$  zeigt eine massive Abweichung von der Homogenität an. Das Zentrum 4 stellt 6mal die Indikation "Mischdiagnose", die in keinem anderen Zentrum der Studie sonst auftaucht. Zur Prüfung der Elimination wird entsprechend unserer Empfehlung jedes einzelne Zentrum den gepoolten anderen Zentren in einer  $2 \times 4$  - Kontingenztafel gegenübergestellt. Es ergeben sich folgende  $X^2$ , die jeweils nach 3 Freiheitsgraden zu beurteilen sind:  $X^2(Z1) = 6,037$ ,  $X^2(Z2) = 3,106$ ,  $X^2(Z3) = 7,958$  und  $X^2(Z4) = 33,114$ . Wie man sieht, ist vor allem Z4 für die Heterogenität verantwortlich zu machen und muß deshalb eliminiert werden.

### 3.3. Das Interaktionskriterium

Es bleibt noch die eingangs aufgestellte Forderung 3 zu kontrollieren:

#### 3. Überprüfe, ob die Zentren bezüglich der Wechselwirkungen in den Faktoren homogen sind.

In den Test werden nur die Wechselwirkungen einbezogen, die in der Auswertung überprüft werden (in unserem Beispiel sind das Medikation  $\times$  Geschlecht und Medikation  $\times$  Indikation). Man bilde für jedes Zentrum die Kreuzklassifikationstabellen aus den zu prüfenden Faktoren und berechne die entsprechenden  $X^2$ -Werte. Die Summe dieser (unabhängigen)  $X^2$  über alle Zentren ist nach  $Z$  Freiheitsgraden zu bewerten. Wird dieser Test signifikant, können simultane Einzeltests pro Zentrum nachgeschoben werden, in denen jeweils die Kreuzklassifikationstabellen eines Zentrums den gepoolten Tabellen der anderen Zentren gegenübergestellt werden. Zur Demonstration des Vorgehens berechnen wir aus unserem Beispiel die Homogenität für die Wechselwirkung Geschlecht  $\times$  Medikation (Tabelle 3).

Die Vierfelder- $X^2$ -Werte der einzelnen Zentren sind:  $X^2(Z1) = 7,00$ ,  $X^2(Z2) = X^2(Z4) = 0,40$  und  $X^2(Z3) = 0,06$ , so daß sich eine Summe von  $X^2 = 7,86$  ergibt, die nach  $df = 4$  zu beurteilen ist. Da dieser Wert über einer Signifikanzschranke von  $\alpha = 5\%$  liegt, wird von Homogenität der Interaktion ausgegangen.

		MÄN	WEI	I	SUM
Z1	DIA	0	4	I	4
	PRA	3	0	I	3
	SUM	3	4	I	7
Z2	DIA	2	3	I	5
	PRA	3	2	I	5
	SUM	5	5	I	10
Z3	DIA	6	10	I	16
	PRA	5	10	I	15
	SUM	11	20	I	31
Z4	DIA	2	3	I	5
	PRA	3	2	I	5
	SUM	5	5	I	10

Tabelle 3: Die Kreuzklassifikation der Faktoren Geschlecht (MÄN - WEI) und Behandlung (DIA - PRA) innerhalb der vier Zentren Z1 - Z4.

Nur wenn man von vornherein hypostasiert, daß in Z1 gemäß dem Geschlecht und nicht plangerecht alloziert wurde, kann man die Vierfeldertafel des Z1 mit  $(a,b,c,d) = (0,3,4,0)$  den kombinierten anderen Vierfeldertafeln mit  $(a,b,c,d) = (2+6+2 = 10, 3+10+3 = 16, 3+5+3 = 11, 2+10+2 = 14)$  gegenüberstellen. Die resultierenden Vierfelder- $X^2$ -Werte errechnen sich zu  $X^2(Z1) = 7.00$  und  $X^2(Z2+Z3+Z4) = 0.16$ , sodaß das Gesamt- $X^2 = 7.16$  mit  $df = 1+1 = 2$  Freiheitsgraden auf dem 5%-Niveau signifikant wird und damit Z1 zu eliminieren wäre.



Damit ergibt sich als Ergebnis der Homogenitätsprüfung, daß nur Z4 wegen seiner Indikationsstellung (Prüfung nach Forderung 2) aus der Studie zu eliminieren ist. Zentrum 1 kann trotz deutlicher Abweichung von der Wechselwirkungshomogenität in der Studie belassen werden, wenn nicht bereits vor der Auswertung Bedenken über die plangerechte Allokation in diesem Zentrum bestanden haben.

#### 4. Diskussion

Die zur Homogenitätsbeurteilung multizentrischer Studien empfohlenen 3 Tests sind durchweg heurostatistischer Natur und beanspruchen nicht, das vereinbarte  $\alpha$ -Risiko der statistischen Entscheidung auch einzuhalten, wie dies bei inferenzstatistischen Prozeduren gefordert wird. Da es jedoch jeweils um die Beibehaltung einer  $H_0$  und nicht um deren Ablehnung geht (wie beim üblichen Testen auf Signifikanz), ist die Einhaltung von  $\alpha$  garnicht wünschenswert, weil andernfalls das beta-Risiko unverhältnismäßig hoch ansteigt und tatsächliche Heterogenitäten nicht nachzuweisen sind.

Dieselbe Argumentation wie für die 3 simultan durchzuführenden Tests gilt für die im Fall der Inhomogenität empfohlenen Eliminationstests. Auch sie sind nur unter heuristischem Aspekt zu beurteilen, zumal eine angemessene Alpha-Adjustierung für eine zweistufige Testprozedur (wie sie aus den nachgeschobenen Zentrentests resultiert) noch nicht existiert (siehe dazu KRAUTH, 1980).

Klinische Populationen sind ex natura meist extrem heterogene Populationen von Ptn. Es ist daher nicht zu erwarten, daß Behandlungszentren Zufallsstichproben aus solchen Populationen sind. Vielmehr muß damit gerechnet werden, daß die Untersuchungszentren ihrerseits Stichproben aus je besonderen Populationen sind. Aus diesen klinischen Erwägungen scheint es nicht sinnvoll, den Begriff der Zentren-Homogenität eng zu definieren und etwa zu fordern, daß alle Z Zentren bezüglich bestimmter Ausgangswerte homogen erscheinen.

Das hier vorgeschlagene Vorgehen zur Homogenitätsbeurteilung geht von dem Grundgedanken aus, daß die Homogenität von Zentren nur an jenen faktoriellen Variablen überprüft wird, die später in die Auswertung eingehen. Damit ist es weniger streng als Ansätze, die alle erhobenen Variablen heranziehen. Andererseits ist das Vorgehen strenger, wenn es zusätzlich die Homogenität der Wechselwirkungen der faktoriellen Variablen verlangt. Das geschieht aus dem Gedanken, daß die Wechselwirkung im Sinne der differentiellen Wirkung von Behandlungen zu interpretieren ist. Weiter führt der Verzicht auf eine  $\alpha$ -Adjustierung zu einem strengeren Vorgehen. Dies ist allerdings zu vertreten, da die Zahl der Tests relativ gering ist.

#### Literaturverzeichnis

- CAMPBELL, D.T. & STANLEY, J.C. (1963). Experimental and quasi-experimental designs for research in teaching. In: N.L. GAGE (Ed.): Handbook of research in teaching. Chicago: Rand McNally.
- KRAUTH, J. (1980). Ein Vergleich der Konfigurationsfrequenzanalyse mit der Methode der log-linearen Modelle. Zeitschrift für Sozialpsychologie 11, 233-247
- LIENERT, G.A. & KRAUTH, J. (1979). Die Konfigurationsfrequenzanalyse. Teil IX. Zeitschrift für Klinische Psychologie und Psychotherapie 22, 3-17

Adresse der Autoren:

Dr.R. Kohnen, PD Dr.H.-P. Krüger, Prof.Dr. G.A. Lienert  
Regensburger Str.160  
D 8500 Nürnberg (West Germany)