

Franz PETERMANN¹/Franz-Josef HEHL²/
Wolfgang SCHNEIDER³:

Veränderungsmessung im Rahmen der klassischen Testtheorie

Der Beitrag diskutiert die Probleme der Veränderungsmessung im Rahmen der klassischen Testtheorie.

Ausgehend von den Axiomen der klassischen Testtheorie (vgl. LORD und NOVICK 1968) werden exemplarisch für den klinischen Bereich verschiedene Veränderungsindices abgeleitet. Diese werden zunächst vergleichend theoretisch diskutiert. In die Darstellung fließen dann herkömmliche Auswertungsstrategien der Veränderungsmessung ein (nonparametrische Tests, Pfadanalyse, Cross-Lagged-Analysis, multivariate Varianzanalyse, Faktorenanalyse etc.).

Eine Extension der klassischen Testtheorie stellt der Ansatz von BEREITER (1963), die Einführung änderungssensitiver Items, dar. – Alle methodischen Erwägungen sollen möglichst an Forschungsbeispielen der Klinischen Psychologie erörtert werden. Abschließend sollen für die Forschungspraxis Handlungskonzepte erkennbar sein.

Einleitung

Wir wollen uns an dieser Stelle des Bandes zum ersten Mal umfassend mit den Problemen der Veränderungsmessung befassen. Diese Probleme sind für die Klinische Forschung von so eminenter Bedeutung, daß ihnen ein ganzes Kapitel des vorliegenden Handbuchs gewidmet ist. Als Grundfrage geht in alle Darstellungen ein:

WIE KOMME ICH ZU MESSWERTEN, DIE EINE STATTGEFUNDENE VERÄNDERUNG WIDERSPIEGELN, UND WIE KANN MAN SIE INTERPRETIEREN?

Diese Frage wird in jeder Untersuchung direkt oder indirekt beantwortet, und dies spätestens an der Stelle, an der ein entsprechender Veränderungsindex zur Beschreibung der Veränderung eingeführt wird. Diese Veränderungsindices können nach BAYER (1974) folgenden Zielen dienen:

1, 3 Psychologisches Institut der Universität Heidelberg.

2 Klinisches Institut für Sozial- und Arbeitsmedizin der Universität Heidelberg.

1. „dem Effektnachweis, d.h. dem Nachweis, daß überhaupt eine Veränderung stattgefunden hat;
2. der Effektdetermination, d.h. der Frage, durch welche Variable eine Veränderung hervorgerufen wird (Aufklärung der Veränderungsvarianz) und
3. der Bestimmung der Effektstruktur, d.h. in welchem Ausmaß die effektbedingten Veränderungen die Beziehungen der Variablen untereinander beeinflussen“ (BAYER 1974, S. 166f.).

CATTELL (1966a) stellt in einer Übersichtsarbeit anhand einer Zufallsauswahl von 100 Zeitschriftenartikeln fest, daß über 50% von ihnen sich mit Veränderungsmessung befassen. Dieses Interesse an der Veränderungsmessung wird von allen klinischen Forschungsbereichen in ähnlich starker Weise bekundet (vgl. FISKE et al. 1970, MCNAMARA und McDONOUGH 1972, alle in diesem Band; FAHRENBURG 1968, BASTINE 1970, HARTIG 1975 u. v. a.). Das starke Interesse bringt es mit sich, daß in verschiedenen klinischen Bereichen (Physiologie, Psychophysiologie, klinische Therapieforschung, experimentelle Psychologie etc.) aus dem unterschiedlichen Kontext der Forschung her jeweils andere Fragestellungen in den Vordergrund rücken. Die angedeutete Vielfalt und Unterschiedlichkeit drückt sich am klarsten in den Begriffen aus, die in den einzelnen Disziplinen ihren Einzug fanden (vgl. PETERMANN 1975). Hier treten unter anderem folgende Begriffe auf:

- Verlaufsforschung (FAHRENBURG 1967, 1968, 1969; FAHRENBURG und MYRTEK 1967, RENN 1974);
- Prozeß- und Erfolgsvorschung (vgl. BASTINE 1970, HARTIG 1975);
- Mehrzeitpunktuntersuchungen (MAYER und PETERMANN 1976);
- Zeitreihenexperiment/Zeitreihenanalyse (CAMPBELL 1957, 1959, 1963, CAMPBELL und STANLEY 1969, FAHRENBURG 1968);
- Panelstudie (NEHNERASJA 1974).

Als der unserer Meinung nach neutralste Begriff bietet sich der der MEHRZEITPUNKTANALYSE/MEHRZEITPUNKTUNTERSUCHUNG an (vgl. auch KNOPF und PETERMANN 1976, PETERMANN und KNOPF 1976).

1. Annahmen der klassischen Testtheorie

Die Diskussion der Veränderungsmessung erstreckte sich in den letzten Jahren vor allem auf die klassische Testtheorie. Diese geht von einer deterministischen Beziehung zwischen gemessenen Größen (x), wahren Größen (X) und dem Fehler der Messung (e_x) aus. Es handelt sich hierbei um einen linearen Ansatz: $x = X + e_x$.

Davon abgesetzt wird die probabilistische Testtheorie, die auf RASCH (1960) zurückgeht (vgl. HEHL und PETERMANN in diesem Band).

Als wichtige Zusammenfassung des Forschungsstandes zur klassischen Testtheorie kann die Arbeit von O'CONNOR (1972) angeführt werden. Nach den Modellvorstellungen der klassischen Testtheorie lassen sich folgende Axiome aufstellen (vgl. LORD und NOVICK 1968, LIENERT 1969, FISCHER 1974):

1. ein beobachteter Wert x_i setzt sich zusammen aus dem wahren Merkmalsanteil X_i und einem Fehleranteil e_i ;
2. der durchschnittliche Meßfehler in der beliebigen Population ist Null;
3. der Meßfehler e_i ist mit dem wahren Wert X_i unkorreliert und
4. Meßfehler aus verschiedenen Tests sind unkorreliert, sofern die Test experimentell unabhängig vorgegeben werden – d.h. lokale stochastische Unabhängigkeit aufweisen (vgl. HEHL und PETERMANN in diesem Band).

Auf die Veränderungsmessung übertragen, ergeben sich aus den vier Axiomen einige Konsequenzen, die nachfolgend diskutiert werden sollen. Zur Erläuterung der Diskussion und aus didaktischen Überlegungen sollen zunächst die verwendeten Notationen zusammengestellt werden (vgl. Tabelle 1).

Tab. 1: Notationen in der klassischen Testtheorie

Messung bei Therapiebeginn: $x = X + e_x$
x = Wert, der beim Therapiebeginn gemessen wird X = wahrer Wert beim Therapiebeginn e_x = Meßfehler von x
Messung nach Abschluß der Therapie: $y = Y + e_y$
y = Wert, der nach Abschluß der Therapie gemessen wird Y = wahrer Wert nach Abschluß der Therapie e_y = Meßfehler von y
Veränderung während der Therapie
g = $y - x$; beobachtete Veränderung G = $Y - X$; wahre Veränderung
Zusammenhänge nach der klassischen Testtheorie
$x = X + e_x$ oder $X = x - e_x$ $y = Y + e_y$ oder $Y = y - e_y$ $g = Y - X + e_y - e_x$ oder $g = G + e_y - e_x$

2. Artefakte in der Veränderungsmessung

Hier sollen alle diejenigen Fehler diskutiert werden, die auftreten können, wenn von einer experimentellen Kontrolle und/oder statistischen Korrektur im Rahmen der Veränderungsmessung abgesehen wird.

Ausgeklammert werden hierbei allgemeine Störfaktoren wie Testungseffekte, selektive Ausgangsstichproben („Patienten“), Drop-Outs, Fehler des Meßinstruments etc. Diese Aspekte werden von KIRCHNER, KISSEL, PETERMANN und BÖTTGER (1976) im Rahmen der internen und externen Validität von klinischen Studien aufgegriffen. Das hier zu diskutierende Grundproblem liegt einerseits in der Meßfehlerbelastbarkeit der Einzelmessungen (x , y) und andererseits in der Korreliertheit der Meßfehler untereinander, die dann auftritt, wenn statistische Abhängigkeiten der Differenz (g) von ihrem Ausgangs- und Endwert nachweisbar sind. Geht man von der Auffassung BEREITERS (1963) aus, lassen sich für die Veränderungsmessung die folgenden drei Problembereiche postulieren:

- Regressionseffekte,
- Reliabilitäts-Validitäts-Dilemma,
- „Physicalism-Subjectivism-Dilemma (Schwierigkeiten bei der Interpretation von Differenzwerten).

Diese Probleme sollen nachfolgend ausführlich diskutiert werden.

Regressionseffekte

Um keine Verwirrung entstehen zu lassen, soll zuerst darauf hingewiesen werden, daß es neben dem statistischen Regressionseffekt, einem Artefakt also, die natürliche Regression zur Mitte gibt. Als natürliche Regression kann man demnach die Wahrscheinlichkeit bezeichnen, daß sich Meßwerte dem Populationsmittelwert annähern, anstatt extremer zu werden.

Beispiel:

Eine zu therapierende Gruppe hat einen höheren Wert auf der Neurotizismus-Skala als die Wartegruppe. Allein aufgrund der natürlichen Regression muß die Therapiegruppe größere Veränderungen in Richtung des Populationsmittelwerts aufweisen als die Wartegruppe. Und dies auch ohne Therapie!

Umgekehrt ist auch der Fall möglich, daß Therapieeffekte unterschätzt werden.

Bei der statistischen Regression sind im Fall von extremen Ausgangswerten die Richtungen der Zufallsänderungen eindeutig determiniert. Diese „Richtungs-determination des Fehlers“ ist in der Therapieforschung von großer Bedeutung, da Klienten/Patienten-Populationen in extremen Bereichen angesiedelt sind. So müssen etwa EYSENCKS „Spontane Remissionen“ unter der Berücksichtigung der natürlichen und statistischen Regressionseffekte gesehen und interpretiert werden (vgl. hierzu auch KIESLER 1976).

Beispiel:

Zur Verdeutlichung des statistischen Regressionseffektes soll ein hypothetisches Beispiel konstruiert werden:

Vor der Zuweisung zur Einzeltherapie werden bei einer Gruppe von Erwachsenen die Neurotizismus-Werte nach EYSENCK bestimmt. Als Meßinstrument dient ein herkömmlicher Fragebogen. Nach der Therapie wird eine Zweitmessung durchgeführt (siehe hierzu Abb. 1).

Bei der gegebenen Fragestellung wird man besonders Gewicht auf die jeweiligen Extremgruppen der „Veränderer“ legen. Man könnte beispielsweise lediglich diejenigen Patienten herausgreifen, die sich um mehr als fünf Punktwerte verbesserten oder verschlechterten. Durch diese Betrachtung der Extremgruppen wird der Regressionseffekt besonders deutlich: Patienten mit hohen Neurotizismus-Werten in der ersten Messung erlangen in der zweiten niedrigere und umgekehrt. Der Regressionseffekt wird besonders häufig auch in der Legasthenieforschung provoziert, wenn zu Übungszwecken besonders gute und besonders schlechte Leser nach der Methode der „matched pairs“ zusammengestellt werden. Die bessere Leistung der Legastheniker in einem Nachtest wird oft unkritisch auf das spezifische Übungsprogramm zurückgeführt.

Der Regressionseffekt hat seine Ursache darin, daß Messungen fehlerbelastet sind und somit mit einer gewissen Wahrscheinlichkeit der Fall eintritt, daß Kennwerte (z.B. Neurotizismus-Werte) um einen bestimmten Betrag über- bzw. unterschätzt werden. Führt man mehrere Messungen an einer Person durch, ist es unwahrscheinlich, daß der Fehler immer gleich bleibt. Vielmehr wird man folgendes beobachten:

+ Patient (1) besitzt einen niedrigen Meßwert, wobei es sehr wahrscheinlich ist, daß der Meßfehler nach dem Zufall den niedrigen Meßwert mitverursacht. Bei der

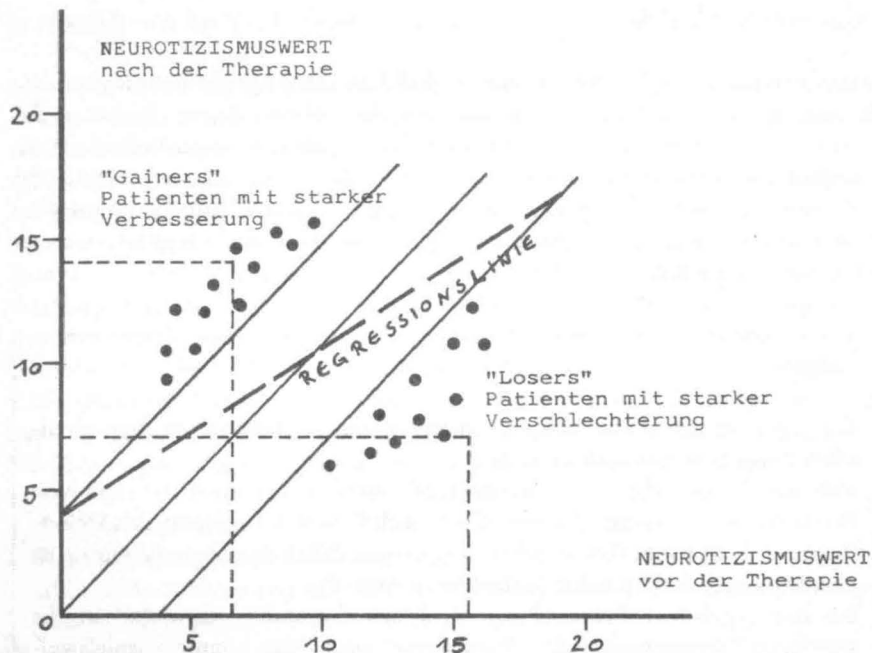


Abb. 1: Graphische Darstellung des Regressionseffekts

folgenden Messung wird Patient (1) seinem wahren Wert näherkommen oder sogar überschätzt werden. Es ergeben sich die Effekte in Abb. 2.

+ Patient (2) besitzt einen hohen Meßwert, wobei es sehr wahrscheinlich ist, daß der Meßfehler nach dem Zufall den hohen Meßwert mit verursacht. Bei der folgenden Messung wird Patient (2) seinem wahren Wert näherkommen oder sogar unterschätzt werden. Es ergeben sich die Effekte in Abb. 2.

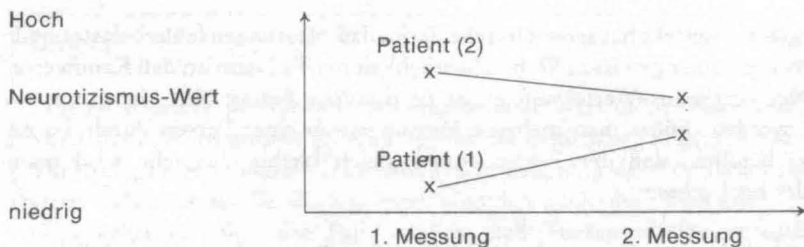


Abb. 2: Regression zur Mitte

Nachzutragen ist, daß der Regressionseffekt sich dann besonders verhängnisvoll auswirkt, wenn neben den Variablen x und y noch eine dritte Variable w (z. B. Angst) eingeführt wird, die mit der Veränderung des Neurotizismus-Wertes in Zusammenhang steht. Führt man jetzt aufgrund der Rohwertdifferenzen eine statistische Analyse durch, werden die erzielten Ergebnisse durch den Regressionseffekt entscheidend beeinflusst, denn die zu untersuchenden Veränderungsbeträge hängen von den jeweiligen Ausgangswerten ab. Korrekturverfahren werden unter Punkt 4. diskutiert.

Reliabilitäts-Validitäts-Dilemma

Innerhalb der klassischen Testtheorie kann man nicht ohne weiteres von der beobachteten auf die wahre Struktur schließen; tut man dies dennoch, müssen Angaben über die „Güte“ der jeweiligen Aussage gemacht werden. Die zu diskutierende Fragestellung lautet also: wie schließt der Kliniker von dem gemessenen Therapieeffekt $-g = y - x -$ auf den wahren Effekt seiner Therapie $-G = Y - X -$, und welche Probleme ergeben sich dadurch?

Hierzu ist es nötig, die Reliabilität der Differenzen zu bestimmen, denn es muß festgestellt werden, ob die gemessene Veränderung g auf eine Verhaltensfluktuation (FAHRENBERG und MYRTEK 1967) oder eine Zufallsfluktuation zurückzuführen ist. Stellt sich die Differenz als eine Zufallsfluktuation heraus, so ist sie auf den Meßfehler zurückzuführen ($e_x = x - X$).

Bestimmung der Reliabilität der Differenzen

Wir gehen zunächst von der Bestimmung der Reliabilität der Einzelmessungen aus:

$$r_{xx'} = 1 - \frac{s_x^2}{S_x^2}$$

$$r_{yy'} = 1 - \frac{s_y^2}{S_y^2}$$

wobei

s_x^2 bzw. s_y^2 die wahre Varianz und

S_x^2 bzw. S_y^2 die Gesamtvarianz darstellen.

In Anlehnung an LORD (1963, S. 28ff.) ergibt sich dann für die Reliabilität der Differenzen folgender Ausdruck:

$$r_{gg'} = \frac{s_y^2 r_{yy'} - 2 s_{yx} r_{xy} + s_x^2 r_{xx'}}{s_y^2 - 2 s_{yx} r_{xy} + s_x^2}$$

wobei

s_{yx} die Kovarianz zwischen y und x und

r_{xy} die Validität des Prätests für den Posttest darstellt.

Bei z-transformierten Variablen ($s_x^2 = s_y^2$) erhält man folgende Vereinfachung des Ausdrucks:

$$r_{gg'} = \frac{r_{yy'} - 2 r_{xy} + r_{xx'}}{2(1 - r_{xy})}$$

Das Reliabilitäts-Validitäts-Dilemma kann man wie folgt umschreiben:

1. Je höher die Korrelation zwischen der 1. und 2. Messung, desto niedriger die Reliabilität der Differenzwerte; je mehr die 1. Messung der 2. aber ähnelt, desto weniger benötigt man die Differenzwerte.
2. Je niedriger die Korrelation zwischen der 1. und 2. Messung, desto unwahrscheinlicher ist es, daß der Test zu beiden Zeitpunkten das Gleiche mißt.

Beispiel:

Greifen wir wieder auf unser Neurotizismus-Beispiel zurück. Für das Beispiel sei $r_{xx'} = 0,83$ und $r_{yy'} = 0,85$; der Wert von r_{xy} liege bei 0,83.

Setzt man diese Werte in die Gleichung zur Bestimmung der Reliabilität der Differenzwerte ein, erhält man:

$$r_{gg'} = \frac{0,83 - 1,66 + 0,85}{2(1 - 0,83)}$$

Der Wert für $r_{gg'}$ beträgt 0,06, die Reliabilität liegt demnach nahe Null.

GEDANKENEXPERIMENT: Zur Verdeutlichung des Problems kann man ein Gedankenexperiment durchführen und prüfen, wie hoch die Reliabilität der Einzelmessungen sein muß, um brauchbare Reliabilitätsschätzungen für die Differenzwerte zu erhalten. Um das Ergebnis vorwegzunehmen: man müßte die Reliabilität bei mindestens 0,95 ansetzen, um solide Reliabilitätsschätzungen zu bekommen.

„Physicalism-Subjectivism-Dilemma“ (Schwierigkeiten bei der Interpretation von Differenzen)

Das zentrale Problem bei der Interpretation von Veränderungswerten wird von BEREITER (1963) als das „physicalism-subjectivism-dilemma“ umschrieben. Dieser Begriff deutet auf eine Problematik hin, die dann auftaucht, wenn exakt quantifizierbare Merkmalsbereiche erschlossen werden sollen, die zugleich psychologisch interpretierbar sind.

Beispiel:

Ein berufserfahrener Therapeut bekommt häufiger „schwierigere“ Patienten zugewiesen als ein weniger erfahrener. Schon aufgrund des höheren Ausgangsniveaus dieser Patienten kann der erfahrene Therapeut höhere Veränderungswerte evozieren. Die erzielten Veränderungsbeträge lassen sich jedoch nicht ohne weiteres in qualitativer Hinsicht miteinander vergleichen.

Gehen wir im Beispiel zusätzlich davon aus, daß während der Therapiesitzungen bestimmte Persönlichkeits- und Verhaltensvariablen des Therapeuten erhoben werden. Nun tritt folgender Fall ein: die erhobenen Variablen korrelieren mit den Veränderungswerten und werden in der Folge als therapiebegünstigend interpretiert. Vor solch vorschnellen Entscheidungen soll hier gewarnt werden.

4. Veränderungsmaße im Rahmen der klassischen Testtheorie

Veränderungsmessung ist solange problemlos möglich, wie die Fehlerquellen der Messung kontrolliert werden können. Wenn dies der Fall ist, kann man mit Differenzänderungswerten arbeiten ($d_i = x_i - y_i$). Eine solche Erfassung der „Veränderung“ scheint in der klinisch-psychologischen Forschung wenig realistisch zu sein. Deshalb sollen hier einige Indices vorgestellt und diskutiert werden, die Veränderung anzeigen können.

Es werden

- einfache Differenzwerte,
- Korrelationsmaße,
- Regressionsmaße und
- Residualmaße

behandelt. Die Ergebnisse der Diskussion sind in einer anschließenden ÜBERSICHTSTABELLE dargestellt.

Einfache Differenzwerte

Man geht von zwei Messungen aus und vergleicht den Anfangs- und Endwert der Messung ($d_i = x_i - y_i$). Dieses Maß wurde u. a. von BASTINE (1970) und FAHRENBURG (1968) kritisiert, da es zu Fehleinschätzungen der Änderungsbeträge für die Personen führt, die in der Ausgangsmessung extreme Werte aufweisen. Diese Effekte kann man durch die Regression der Meßwerte zur Mitte (s. o.: statistische Regression) erklären. Wir haben oben darauf hingewiesen, daß hohe Werte zu niedrigen und niedrige zu höheren regredieren.

Korrelationsmaße

Ein einfaches Korrelationsmaß stellt die Korrelation zwischen der Ausgangsmessung x_i und dem Differenzwert d_i dar. Die Schwächen dieses Maßes liegen in der Tatsache begründet, daß weder die Meßfehler der Einzelmessungen (x_i und y_i) noch die Abhängigkeit der Messungen untereinander korrigiert werden. Dies hat besonders dann verblüffende Folgen, wenn die Fehlerkomponenten in unterschiedlicher Weise und Richtung wirken; eine Folge wird z. B. sein, daß die Korrelation (r_{xd}) zu niedrig ausfällt. Da die meisten Veränderungsindices auf diese Weise berechnet wurden (vgl. BEREITER 1963), müssen die damit erzielten klinischen Forschungsergebnisse sehr vorsichtig interpretiert werden.

Regressionsmaße

Wie oben erwähnt, liegt die Schwäche der Differenzmaße u. a. in der Fehlerhaftigkeit der Einzelmessungen. Eine Möglichkeit besteht nun darin, die Einzelmessungen x und y über Regressionsschätzungen zu korrigieren; man erhält $D = Y - X$. Somit ist zwar ein Nachteil – die Fehlerhaftigkeit der Einzelmessungen – ausgeschaltet, jedoch berücksichtigt dieser Ansatz nicht die Abhängigkeit der Messungen untereinander. LORD (1958, 1963) erkannte dies und schlug ein Schätzverfahren vor, das x und y über Partialkorrelationen bereinigt in die Schätzgleichung mit einbezieht. CRONBACH und FURBY (1970) kommen in einer Übersichtsarbeit zu dem Ergebnis, daß unter den bisher vorgestellten Korrekturverfahren das LORD-Verfahren am günstigsten abschneidet.

CRONBACH und FURBY erweiterten diesen Ansatz jedoch in der Weise, daß sie noch eine Korrelation der Veränderungsbeträge mit den Drittvariablen vornah-

men. Konkret schlugen sie vor, zwei Außenvariablen – für die Therapieforschung könnten dies u.U. die Therapiedauer, die Persönlichkeits- und Verhaltensvariablen des Therapeuten/Patienten u.ä. sein – in die Regressionsschätzung einzubeziehen. Die Schätzformel für den Veränderungsbetrag sieht dann wie folgt aus:

$$\hat{D} = b_1x + b_2y + b_3z + b_4w.$$

Die Gewichtungszahlen b_1 bis b_4 werden so bestimmt, daß die Differenz ($\hat{D} = D$) ein Minimum wird. Die Korreliertheit der Einzelmessungen wird über die Partialkorrelation ausgeschaltet.

Der Ansatz von CRONBACH und FURBY ist bei weitem der exakteste zur Schätzung der Veränderungseffekte; es können zudem noch die Effekte der Außenvariablen auf die Veränderung geschätzt werden. Schwierigkeiten ergeben sich jedoch dadurch, daß oft nicht von vornherein klar ist, welche Außenvariablen für den therapeutischen Prozeß wichtig sind und in die Schätzgleichung einbezogen werden müssen.

Residualmaße

Bei der Verwendung von Differenzwerten tritt oft nur eine geringe Korrelation mit relevanten Dimensionen des Veränderungsprozesses auf. Hier schlug LORD (1958, 1963) die Verwendung von Residualmaßen vor, die den Vortestwert (x) aus dem Nachtestwert (y) herauspartialisieren. Doch auch bei der Benutzung des Residualmaßes treten Verzerrungen auf, die BEREITER (1963) und KNOPF und PETERMANN (1976) diskutieren.

Eine Weiterentwicklung des einfachen Residualansatzes stellt der Ansatz von TUCKER, DAMARIN und MESSICK (1966) – base-free measure of change – dar, „indem das wahre X , errechnet über die Reliabilitätskorrektur, und nicht aber das fehlerhafte x aus der Nachmessung y herauspartialisiert wird“ (KNOPF und PETERMANN 1976). O’CONNOR (1972) kritisiert den Ansatz, da er sich aus einem unkorrigierten y und einem fehlerkorrigierten X zusammensetzt. CRONBACH und FURBY modifizierten ihn in der schon oben diskutierten Weise, nämlich durch die Einbeziehung von Drittvariablen. Die Schätzung der Veränderung erfolgt danach über die sukzessive Auspartialisierung der korrigierten Variablen \hat{X} und \hat{W} . Die Schätzgleichung lautet dann:

$$\hat{Y} \cdot \hat{X}\hat{W} = b_1x + b_2y + b_3w + b_4z.$$

Hier stellt sich allerdings ebenfalls wieder die Frage nach der Auswahl der relevanten Drittvariablen.

Resümee

Insgesamt gesehen scheint kein adäquates Veränderungsmaß vorzuliegen. Jedoch ist es falsch, an dieser Stelle zu resignieren, wie dies viele Autoren zu tun pflegen. Es scheint evident, daß die zuletzt diskutierten Maße trotz einer nur partiellen Lösung des Problems einige Vorteile bieten (z.B. die Einbeziehung der Drittvariablen). Verhängnisvoll erscheint dagegen eine Überlegung von RICHARDS (1975) und RICHARDS et al. (1975). Der Autor versucht, mit Hilfe eines undurchsichtigen Computerexperiments, das in der Aussage mündet: „alle Veränderungsindices sind gleich, nehmen wir halt das klassische Differenzmaß, da

Tab. 2: Zusammenstellung der wichtigsten Veränderungsindices

VERÄNDERUNGSINDICES	
VERÄNDERUNGSINDEX	BEMERKUNGEN
einfache Differenzwerte	einfache Differenzwertbildung nach $d_i = x_i - y_i$. <i>Vorteil:</i> – ökonomisch – populationsunabhängig. <i>Nachteil:</i> – Regressionseffekt, – Differenzen erfordern mindestens Intervallskalenniveau.
Korrelationsmaße	einfache Korrelation zwischen Anfangswert x_i und Änderungswert d_i . <i>Vorteil:</i> ökonomisch. <i>Nachteil:</i> – die Meßfehler der Einzelmessungen werden nicht korrigiert; – die serielle Abhängigkeit der Messungen untereinander wird nicht berücksichtigt.
Regressionsmaße	Die Fehlerhaftigkeit der Einzelmessungen wird über Regressions-schätzungen korrigiert; in der Praxis kommt hierzu die Kovarianzanalyse zum Einsatz. Die Kovarianzanalyse bereinigt z.B. in einer Zweizeitpunktmesung den Wert der zweiten Messung aufgrund des Einflusses der Erstmessung. (Vgl. Diskussion bei RENN (1974), der sich gegen dieses Verfahren ausspricht – u.a. gegen die Annahme der Orthogonalität der Wirkungen.)

Regressionsmaße
(z. B. das LORD-Verfahren)

Vorteil:

- die beiden Einzelmessungen werden über Regressions-schätzungen korrigiert;
- die Abhängigkeiten der Messungen untereinander werden durch Partialkorrelation beseitigt.

Nachteil:

- Drittvariablen bleiben unberücksichtigt.
-

Regressionsmaße
(z. B. das CRONBACH-
FURBY-Verfahren)

Vorteile:

- die Einzelmessungen werden über Regressions-schätzungen korrigiert;
- die Abhängigkeiten werden über Partialkorrelation beseitigt
- Drittvariablen werden einbezogen.

Nachteil:

- es bleibt ungeklärt, warum gerade eine bestimmte Drittvariable einbezogen wird und nicht eine andere.

Zum Verfahren:

Die Schätzformel der Veränderung lautet:

$$\hat{D} = b_1x + b_2y + b_3z + b_4w.$$

Residualmaße (allgemein)

Statt von Differenzmaßen geht man von Residualmaßen aus, bei denen man die Erstmessung (x) aus der Nachmessung (y) herauspartialisiert (vgl. die Diskussion bei BEREITER 1963).

Residualmaße
(z. B. das von
TUCKER et al.)

TUCKER, DAMARIN und MESSICK (1966) versuchten durch ein Konzept des base-free-measure of change das wahre X aus der Nachmessung y herauszupartialisieren.

Bewertung:

Es wird eine unkorrigierte Messung (y) mit einer fehlerkorrigierten (X) verknüpft. Dieser Ansatz wird von O'CONNOR (1972) kritisiert.

Residualmaße
(z. B. das von
CRONBACH und FURBY)

Es werden ebenfalls zwei fehlerkorrigierte Drittvariablen in die Schätzgleichung einbezogen. Die Korreliertheit der Beobachtungen (x, y) wird durch Partialkorrelation berücksichtigt. Die Schätzung der wahren Veränderung ergibt sich aus der sukzessiven Auspartialisierung der fehlerkorrigierten Variablen (Y · XW).

Als Schätzgleichung ergibt sich:

$$\hat{Y} \cdot \hat{X} \hat{W} = b_1x + b_2y + b_3w + b_4z.$$

Bewertung:

Es bleibt hier wieder ungeklärt, warum gerade eine bestimmte Drittvariable einbezogen wird und nicht eine andere.

es so schön ökonomisch ist“, jede Bemühung um ein zuverlässigeres Veränderungsmaß von vorherein zu blockieren.

Vergleichende Übersicht (vgl. Abb. 2)

Die unter Punkt 4 gemachten Ausführungen sollen abschließend noch einmal in einer Gesamtschau dargeboten werden.

5. Auswertungsstrategien

Nonparametrische Tests

Oft können in der klinischen Forschung nur wenige begründete Annahmen über die Daten getroffen werden. In solchen Fällen besteht der einzige Ausweg darin, nonparametrische Verfahren anzuwenden (vgl. FAHRENBERG 1968). Da diese Verfahren in einigen klassischen Lehrbüchern (vgl. SIEGEL 1956, LIENERT 1973) sehr ausführlich beschrieben sind, braucht hier nicht näher darauf eingegangen zu werden.

ÜBERSICHT: NONPARAMETRISCHE TESTS			
SKALEN-NIVEAU \ KENNZEICHEN DER STICHPROBE		ZWEI-ZEIT-PUNKT-MESSUNG	MEHRZEIT-PUNKT-MESSUNG
nonparametrisch	NOMINALSKALA	Vorzeichentest	Q-Test
	ORDINALSKALA	WILCOXON-Test SPEARMAN-Rangkorrelation	FRIEDMAN-Test multiple Vergleiche nach WILCOXON
parametrisch	INTERVALLSKALA	t-Test für abhängige Messungen, Produkt-Moment-Korrelation etc.	Varianzanalyse für abhängige Messungen, etc.

Abb. 3: Auswahl möglicher Testverfahren zur Erfassung der Effekte von Mehrzeitpunktuntersuchungen

Zur Strukturierung der Daten soll eine Übersichtstabelle aufgestellt werden, die die wichtigsten nonparametrischen Tests für abhängige Stichproben enthält (vgl. ähnliche Tabellen in SIEGEL 1956, SACHS 1969). Der Vollständigkeit halber wurden auch parametrische Tests in die Übersicht aufgenommen.

Nachgetragen werden muß, daß der Einsatz nonparametrischen Verfahren nur begrenzte Testmöglichkeiten umfaßt (so treten u.a. Schwierigkeiten bei der Testung von Wechselwirkungen auf; vgl. BREDEKAMP 1974). Methodisch interessierte Leser sollten sich jedoch hierzu die Ausführungen von PURI und SEN (1971) vergegenwärtigen.

Einfache Korrelationsanalyse

Eine Möglichkeit, die Abhängigkeit der Meßwerte nachzuweisen, besteht in der Analyse von Autokorrelationen und serialen Kreuzkorrelationen (vgl. zur Verdeutlichung Abb. 4, nach CATTELL 1966b). Sind die Ergebnisse einer Mehrzeitpunktuntersuchung korreliert, kann man die seriale Korrelation auf diesem Wege nachweisen (vgl. Abb. 4, Fall 1). Die Berechnung basiert auf der Theorie der autoregressiven Prozesse (vgl. ANDERSON 1971, SCHNEEWEISS 1971) und kann bei MAYER und PETERMANN (1976) nachgelesen werden.

Ebenso wie man die Autokorrelation eines Prozesses überprüfen kann, läßt sich dessen Korrelation mit mit einem anderen Prozeß bestimmen (vgl. Abb. 4, Fall 2

Fall (1)	– AUTOKORRELATION (lag = 3)
VAR a	1 2 3 4 5 . n
VAR a	. . . 1 2 3 4 5 (n-3)(n-2)(n-1)n
	(zwischen VAR a und VAR a liegt eine Zeitverschiebung von lag = 3 vor)
Fall (2)	– SERIALKORRELATION
VAR a	1 2 3 4 5 . n
VAR b	1 2 3 4 5 . n
	(zwischen VAR a und VAR b liegt keine Zeitverschiebung vor)
Fall (3)	– SERIALKORRELATION (lag = 2)
VAR a	1 2 3 4 5 . n
VAR b	. . 1 2 3 4 5 (n-2)(n-1)n
	(zwischen VAR a und VAR b liegt eine Zeitverschiebung von lag = 2 vor)

Abb. 4: Darstellung der Serialkorrelation und Autokorrelation (in Anlehnung an CATTELL 1966b)

und 3). Es wird geprüft, ob eine synchrone oder zeitverschobene Kreuzkorrelation vorliegt (vgl. HOLTZMAN 1963, wiederabgedruckt in diesem Band; DAHME 1975, DAHME et al. 1976, DAHME 1976b, in diesem Band). Die Korrelationskoeffizienten werden mit den herkömmlichen Tests auf Signifikanz überprüft (ANDERSON 1971, BOX und JENKINS 1970) und die festgestellten Zusammenhänge in einem Korrelogramm abgetragen.

Somit wird durch einen partiellen serialen Korrelationskoeffizienten bestimmt, in welcher Weise Messungen von vorausgegangenen abhängig sind.

*Cross-Lagged-Analysis*⁴

Es handelt sich um eine quasi-experimentelle Anordnung, um die Richtung eines Kausalzusammenhangs festzustellen. Sie stellt eine Weiterentwicklung der 16-Felder-Tabelle von LAZARSELD (1947) dar. Die von CAMPBELL (1963) vollzogene Erweiterung des Verfahrens machte es möglich, auch nicht-dichotome Variablen zu verarbeiten.

Beispiel:

LAZARSELD'S 16-Felder-Tabelle wird normalerweise dazu benutzt, um die relativen Stärken/Schwächen verschiedener Einstellungen zu erkennen, weiterhin dazu, um die Richtung eines Kausalzusammenhangs zu bestimmen, da die üblichen Korrelationsergebnisse in dieser Hinsicht keinen eindeutigen Schluß zulassen. Angenommen, das Verhalten von 20 Therapeuten würde in einer bestimmten Therapiesituation als „warmherzig“ oder „kühl“ klassifiziert und das Verhalten ihrer Patienten/Klienten als „kooperativ“ und „unbeteiligt“, so könnte sich eine positive Korrelation ergeben: warmherzige Therapeuten haben kooperative Klienten. Es kann nun die Frage gestellt werden, ob die Warmherzigkeit der Therapeuten nun Ursache oder Folge der Mitarbeit der Klienten ist.

Eine Zweipunktmessung könnte weitere relevante Meßwerte liefern, wenn dieselben Variablen bei einer anderen Gelegenheit mit denselben Therapeuten und Klienten erneut untersucht würden (zwei Stufen der Messung ergeben bei zwei Variablen vier Antwortmöglichkeiten für jede Situation oder viermal vier mögliche Schemata für die beiden Situationen, so daß sich eine 16-Felder-Tabelle ergibt). Die Mehrdeutigkeit der üblichen Korrelationsergebnisse zeigt sich, wenn man sich klar macht, daß es zwei verschiedene Möglichkeiten für die Veränderung gibt:

4 Dieser Abschnitt ist leicht modifiziert aus PETERMANN (1976) entnommen.

Die Warmherzigkeit des Therapeuten als Ursache für die Mitarbeit des Klienten

Klienten	Therapeut	
	kühl	warmherzig
kooperativ	2 (2)	↑ 6
unbeteiligt	6 ↓	(2) 2

Die Mitarbeit des Klienten als Ursache für die Warmherzigkeit des Therapeuten

Klienten	Therapeut	
	kühl	warmherzig
kooperativ	2 (2) →	6
unbeteiligt	6 ←	(2) 2

Abb. 5: Cross-Lagged-Analysis

Es wurden nur vier stabile Änderungsformen berücksichtigt, d. h. nur die, die eine Erhöhung der Korrelation zur Folge haben. Unberücksichtigt bleiben die unvermeidlichen zufälligen Veränderungen. Deshalb besitzt obiges Schema (im Gegensatz zu dem von LAZARSELD) nur 8 der 16 Zellen einer vollständigen 16-Felder-Tabelle. Alle vier Verhaltensänderungen können natürlich gleichzeitig auftreten.

Schlußfolgerungen über die Richtung des Kausalzusammenhangs ergeben sich aus der überwiegenden Änderung. Eine solche Form der Analyse, die in der Soziologie vorwiegend auf dichotome Variablen angewendet wird, kann zur Cross-Lagged-Analysis verallgemeinert werden (KENNEY 1975, DUNCAN 1976). Durch diese kann nun die Aussage getroffen werden, daß ein Effekt höher mit der vorangegangenen Ursache als mit einem nachfolgenden Zustand korreliert. Verfahrenstechnisch läßt sich das Cross-Lagged-Modell wie folgt abbilden:

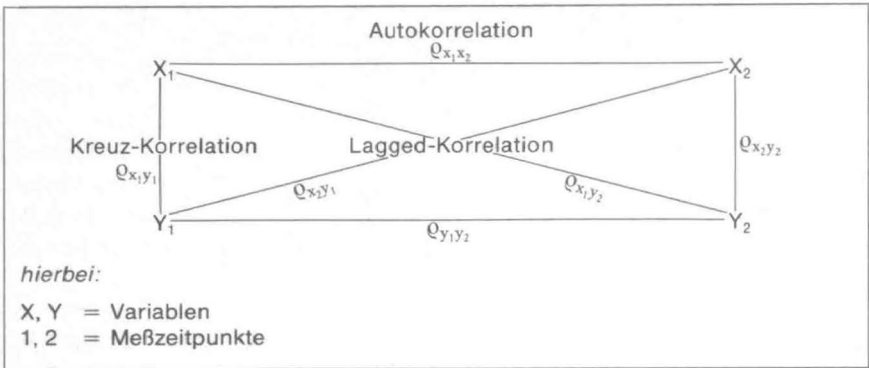


Abb. 6: Cross-Lagged-Modell

Wird Y durch X verursacht, ist die Crossed-Lagged-Differenz ($\rho_{X_1 Y_2} - \rho_{X_2 Y_1}$) positiv. Wirkt der Kausalzusammenhang aber in umgekehrter Richtung, wird die Differenz ($\rho_{X_1 Y_2} - \rho_{X_2 Y_1}$) negativ. Diese stark vereinfachte Darstellung wird von RICHARD (1972), KENNEY (1975) und DUNCAN (1976) präzisiert.

Obwohl die Cross-Lagged-Analysis eine Art Faktorenmodell darstellt, in dem aufgrund der beobachteten Variablen Annahmen über die unbeobachteten Faktoren getroffen werden, lassen sich keine Schätzungen der Faktorenladungen vornehmen. Das Anliegen der Analyse besteht in der MODELLESTUNG und nicht in der Parameterschätzung. Als Modellannahmen werden die Gesetzmäßigkeiten der Veränderungsprozesse (KENNEY 1975) überprüft.

Resümee

Die Stärke der Cross-Lagged-Analysis liegt in der „Spezifität“; sie wurde nämlich speziell für die Mehrzeitpunkt-Analyse⁵ entwickelt (etwa im Gegensatz zur Regressionsanalyse). Es zeigt sich, daß so globale Verfahren wie die Regressions-, Varianz- und Faktorenanalyse für die Fragestellungen einer Mehrzeitpunkt-Analyse viel zu allgemein sind und ihnen somit nur eine beschreibende Funktion zukommen kann.

Es darf nicht übersehen werden, daß der Ansatz alle Schwächen der Quasi-Experimente teilt (CAMPBELL 1963). Trotzdem scheint er der Pfadanalyse (s. u.) in vieler Hinsicht überlegen.

⁵ HEISE (1970) zeigt durch Computersimulation, daß der Meßfehler in Mehrzeitpunkt-Analysen den Regressionskoeffizienten nicht allzu stark verfälscht. Jedoch berücksichtigt diese Simulation nicht den Fall, daß X_1 reliabel und Y_1 weniger reliabel gemessen wurde. Ebenso können der Meßfehler und der Einfluß der Drittvariablen den Ansatz hypothetisch beeinträchtigen.

Abschließend soll noch einmal grundsätzlich auf die Möglichkeiten der Sozialwissenschaften eingegangen werden, Gesetzmäßigkeiten von Ursache und Wirkung zu erkennen. KENNEY (1975, S. 901) vertritt den Standpunkt, daß folgende Schritte eine Überlegung verdienen:

1. "the consistent replication of a cross-sectional relationship;
2. the finding of cross-lagged relationship between cause and effects;
3. the finding of time-lagged differences;
4. an experiment in which the causal variable is manipulated."

KENNEY verdeutlicht, daß die Cross-Lagged-Analysis eine Vermittlerrolle zwischen der Korrelationsanalyse und spezifischen Modellen der Sozialwissenschaften einnehmen kann.

Multivariate Varianzanalyse

BOCK (1963) diskutierte erstmals ein Modell der multivariaten Varianzanalyse für wiederholte Messungen; diese Diskussion erfolgte für den Fall einer univariaten Varianzanalyse für abhängige Messungen insbesondere durch GAITO und WILEY (1963). BOCK zeigte in seiner Arbeit die Einsatzmöglichkeiten der multivariaten Varianzanalyse im Rahmen der Veränderungsmessung auf. MCCALL und APPELBAUM (1973, in diesem Band abgedruckt) wiesen darauf hin, welche Modellannahmen erfüllt sein müssen, um die multivariate (aber auch die univariate) Varianzanalyse erfolgreich in der Veränderungsmessung einzusetzen. Ein sehr ausführliches neueres Lehrbuch zu dieser Fragestellung legte BOCK (1975) vor.

Die Probleme einer multivariaten varianzanalytischen Auswertung sollen an dieser Stelle nicht diskutiert werden; es sei hier auf die Arbeit von MCCALL und APPELBAUM in diesem Band verwiesen.

Pfadanalyse

Die Pfadanalyse ist als nicht-experimentelle multivariate Auswertungsstrategie in neuester Zeit vor allem in der Soziologie aufgegriffen worden (vgl. BLALOCK 1961, 1971, BLALOCK et al. 1968). In der psychologischen Forschung wird der Ansatz u.a. von WERTS und LINN 1970, SEIBEL und NYGREEN 1972 und PEDHAZUR 1975 diskutiert.

Für die Veränderungsmessung in der klinischen Forschung kann dieser Ansatz prinzipiell zur Aufdeckung von Effekten der unabhängigen Variablen auf die abhängigen Variablen (verschiedene Zeitpunkte) herangezogen werden. Die

Pfadanalyse macht hierbei jedoch sehr restriktive Annahmen, die von HEISE (1969), PEDHAZUR (1975), KNOPF und PETERMANN (1976) diskutiert werden. Ohne an dieser Stelle auf die methodischen Problematik des Ansatzes einzugehen, erscheint die Pfadanalyse zur Beschreibung von klinischen Prozessen nicht adäquat zu sein, da es an formalisierbaren klinischen Theorien mangelt.

Kanonische Korrelationsanalyse

Von HARRIS (1963) stammt der Vorschlag, die kanonische Korrelationsanalyse zur Veränderungsmessung einzusetzen (vgl. zum Verfahren selbst GAENSSLEN und SCHUBÖ 1973).

Die herkömmliche Faktorenanalyse geht von einem Variablenset aus, dessen Interkorrelationsmatrix analysiert wird. Bei der kanonischen Korrelation wird dies jedoch anhand zweier Variablensets vorgenommen. Die Faktoren werden jetzt so bestimmt, daß der erste Faktor der Prädiktoren mit dem korrespondierenden ersten Faktor der Kriterien maximal korreliert. Bei den nachfolgenden Faktoren wird dieselbe Vorgehensweise zugrundegelegt. Die Faktoren werden als *kanonische Variablen* bezeichnet, die Korrelationen zwischen einander entsprechenden Faktoren der Prädiktoren- und Kriterienseite als *kanonische Korrelation*.

Die kanonische Korrelation wird in der Veränderungsmessung so angewandt, daß man sich auf die Erst- und Zweitmessung bezieht und diese als getrennte Variablen in der Analyse einführt. Die so „extrahierten Faktoren“ weisen für die Erst- und Zweitmessung jeweils eine Ladung auf. Die Differenz der zugehörigen Ladungszahlen drückt eine *qualitative* Veränderung aus; d.h. inhaltlich ist das in der Erstmessung Erfasste nicht identisch mit dem, was in der Zweitmessung erhoben wurde. Als Beschreibungsdimension der Veränderung dienen die Dimensionen der Faktoren der Ausgangswerte (vgl. HARRIS 1963).

Faktorenanalyse

Ohne auf verfahrenstechnische Fragen eingehen zu wollen (vgl. hierzu CATTELL 1966a, PAWLIK 1968, ÜBERLA 1971), sollen die beiden prinzipiellen Einsatzmöglichkeiten der Faktorenanalyse zur Bestimmung von Veränderungen aufgezeigt werden:

1. Faktorenanalyse getrennt für Werte der Erst- und Zweitmessung,
2. Faktorenanalyse gemeinsam für Werte der Erst- und Zweitmessung.

Zu (1):

Man rechnet zwei getrennte Faktorenanalysen und führt dann beispielsweise eine Ähnlichkeitsrotation nach FISCHER-ROPPERT, KAISER oder KRISTOF durch.

Nachfolgend kann man für jede Person die Faktorenwerte berechnen und erhält somit einen Wert für die Erst- und einen für die Zweitmessung.

Diese Methode, deren Kernstück der Faktorenstrukturvergleich ist, wurde in der Vergangenheit sehr intensiv diskutiert (vgl. SIXTL 1964, FISCHER und ROPPERT 1964, EYFERTH und SIXTL 1965). In der praktischen Anwendung zeigt dieses Verfahren – abgesehen von seiner rechentechnischen Aufwendigkeit – erhebliche Schwächen, die NESSELROADE und BALTES (1970) beschrieben. Die Autoren konnten zeigen, daß eine Wechselwirkung zwischen Stichprobengröße, Variablenanzahl und der Zahl der extrahierten Faktoren bestand. Ebenso sind Fragen der Rotation der Faktorenstruktur und Probleme einer adäquaten Abstimmung der Signifikanz noch nicht ausreichend diskutiert.

Zu (2):

Die Berechnung einer Faktorenanalyse, die sowohl Erst- als auch Zweitmessungen in die Berechnung einbezieht, ist zwar weniger aufwendig, jedoch nur statthaft, wenn man davon ausgehen kann, daß die Faktorenstruktur über die Zeit hinweg stabil bleibt. In die Faktorenanalyse gehen die Werte der Erst- und Zweitmessung dann so ein, als ob es sich um unabhängige Werte handelt, die von zwei verschiedenen Personen stammen. Dadurch verdoppelt sich die Zahl der Versuchspersonen in der Rechnung. Der weitere Prozeß gleicht dem oben beschriebenen (Berechnung von Faktorenwerten pro Person etc.).

Das beschriebene Verfahren ist zwar sehr ökonomisch, jedoch können qualitative Veränderungen in der Weise, daß ein Merkmal eine andere Bedeutung erhält, nicht berücksichtigt werden.

Vergleichende Übersicht

Wie im vorangegangenen Kapitel soll auch hier wieder eine Übersicht über die vorgestellten Verfahren gegeben werden. Die Tabelle ist teilweise in der Form eines Flußdiagramms dargestellt, um die *Vorentscheidungen* zu dokumentieren, die zur Selektion der Verfahren führen. Im Flußdiagramm werden drei Vorentscheidungen zur Diskussion gestellt:

1. eine Vorentscheidung über die Art der Stichprobe,
2. eine Vorentscheidung über die „Art der Methodik“,
3. eine Vorentscheidung über die Art der Verteilungsannahmen.

Sie sollen im folgenden etwas genauer ausgeführt werden.

(1) Vorentscheidung über die Art der Stichprobe

Wir gingen bei der bisherigen Diskussion immer von der Betrachtung von Gruppen bzw. Gruppenmittelwerten aus; wie verschiedene Autoren zeigen (vgl.

HUBER 1973, PETERMANN und KNOPF 1976), muß dies nicht immer vorteilhaft sein. Deshalb bietet sich die Einzelfallbetrachtung als eine weitere Möglichkeit an, um zeitliche Verläufe in der klinischen Forschung zu bestimmen. Dies umso mehr, als in Betrachtungen von Gruppenmittelwerten individuelle Eigenarten der Reaktionen verdeckt werden können (vgl. HYLAND und FOOT 1974). Insofern scheint es angebracht, zeitliche Verläufe bei Einzelpersonen zu ermitteln. Eine detailliertere Diskussion dieser Frage erfolgt an anderer Stelle des Bandes (vgl. HOLTZMAN, DAHME).

(2) Vorentscheidung über die „Art der Methodik“

Der etwas verwirrende Terminus „Art der Methodik“ steht nur symbolisch für die Wahl zwischen konventionellen methodischen Verfahren (beschreibende Modelle; statistische Modelle) einerseits und Verfahren der mathematischen Modellbildung (funktionale Modelle) andererseits. Wir haben hier nur konventionelle methodische Verfahren vorgestellt.

In Zukunft dürften jedoch verstärkt mathematische Modelle zur Anwendung kommen (vgl. FAHRENBERG 1968, MAYER und PETERMANN 1976). Eine Möglichkeit zur Beschreibung von Reaktionsverläufen sieht FAHRENBERG (1968) im Einsatz von Differentialgleichungsmodellen. Ein weniger aufwendig erscheinendes Verfahren bietet die Computersimulation von Veränderungsprozessen. Hierbei besteht die Möglichkeit, theoretische und empirische Ansätze auf ihre Plausibilität hin zu überprüfen, die sich noch im Vorstadium der theoretischen Entwicklung befinden. Eine solche Überprüfung erscheint im Zusammenhang mit den komplexen Fragestellungen von Mehrzeitpunktuntersuchungen auch im klinischen Bereich interessant und durch den Einsatz von Digital- und/oder Analogrechnern ohne großen Programmieraufwand durchführbar zu sein (vgl. HANNAUER 1973).

Hier zeigt es sich, daß durch Computersimulation theoretische und empirische Plausibilitätsannahmen überprüfbar werden.

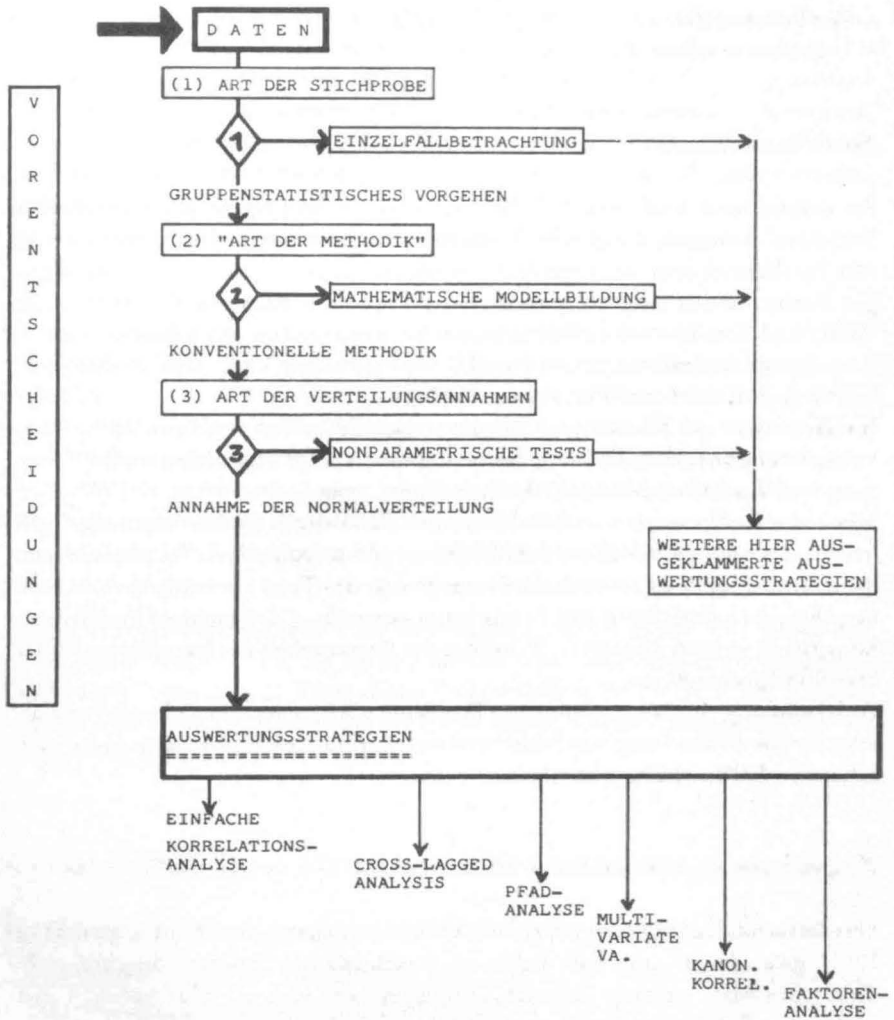
(3) Vorentscheidung über die Art der Verteilungsannahmen

Hierbei geht es um die Entscheidung zwischen parametrischem oder nonparametrischem Vorgehen. Beide Möglichkeiten wurden unter Punkt 5 diskutiert. Beim Vorliegen normalverteilter Daten ergeben sich folgende Auswertungsstrategien:

- einfache Korrelationsanalyse
- Cross-Lagged-Analysis
- Pfadanalyse
- multivariate Varianzanalyse
- kanonische Korrelationsanalyse
- Faktorenanalyse

Die Aufstellung in Tabelle 3 gibt von links nach rechts die steigende Komplexität der Verfahren an. Diese bewirkt, daß auch die Annahmen weniger überprüfbar und damit im klinischen Bereich meist nicht mehr sinnvoll anwendbar sind.

Tab. 3: Auswertungsstrategien



6. Alternativen zur Veränderungsmessung innerhalb der klassischen Testtheorie

Wenn man den Boden der klassischen Testtheorie nicht verlassen will, so bieten sich zwei Strategien an, um ihre Schwächen zu umgehen:

- Paralleltestverfahren
- änderungssensitive Variablen

Paralleltestverfahren

Per definitionem muß bei Paralleltestverfahren davon ausgegangen werden, daß Verfahren vorliegen, die gleiche Merkmalsbereiche erfassen. Durch den Einsatz von Paralleltests sind demnach Meßwertsequenzen mit voneinander unabhängigen Fehleranteilen möglich. Anwendungsbeispiele werden von CATTELL et al. (1961) und THORNDIKE (1966) gegeben. So konstruierten etwa CATTELL et al. einen Persönlichkeitstest in acht Parallelformen, um eine Verlaufsstudie durchzuführen (vgl. KNOFF und PETERMANN 1976).

Im Gegensatz zu KNOFF und PETERMANN, die in ihrem Sammelreferat zur Veränderungsmessung die Paralleltestverfahren als „eine unproblematische Erfassung von Änderungsbeträgen“ kennzeichnen, zeigt jedoch HORST (1971), daß schon die Bestimmung der Reliabilität von Paralleltests große Schwierigkeiten macht: das ermittelte Maß für die Reliabilität gibt sowohl die Reliabilität als auch die zeitliche Stabilität als auch die Homogenität des Tests an. Außerdem ist es in der klinischen Forschung und Praxis kaum vorstellbar, daß mehrere Paralleltests konstruiert werden können (z. B. infolge der Begrenztheit des Itempools und der Spezifitätsproblematik).

Aufgrund der oben angeschnittenen Probleme dürften auch in Zukunft Paralleltests für die Feststellung von fehlerbereinigten Veränderungsbeträgen nur in den seltensten Fällen in Frage kommen.

Änderungssensitive Variablen

Das Konzept der änderungssensitiven Variablen („change items“ nach BEREITER 1963) geht davon aus, daß Items in verschiedenem Ausmaß Effekte, z. B. Therapieeffekte, erfassen können. Erinnern wir uns an das Anliegen der klassischen Testtheorie, das darin bestand, stabile *Merkmale* möglichst genau zu erfassen. Im Gegensatz dazu versucht BEREITER (1963, vgl. auch RENN 1973) ein dynamisches Konstruktionsprinzip einzuführen, das die Änderungssensitivität erfassen will. Danach ist unter Änderungssensitivität der Grad der Regelmäßigkeit

zu verstehen, mit dem ein Testitem bei verschiedenen Personen unterschiedliche Veränderungsbeträge anzeigt.

„Da die Streuung der Meßwertdifferenzen innerhalb einer Stichprobe ($S_d = y - x$) diese Regelmäßigkeit wiedergeben kann, wird die Varianz der Verteilungen der Meßwertdifferenzen (S_d^2) als Kriterium der Änderungssensitivität gewählt. Ein Item ist immer dann änderungssensitiv, wenn $S_d^2 > 0$. Nach einer z-Transformation der beiden beobachteten Meßwertreihen x und y in Meßwertreihen mit gleichem Mittelwert und gleicher Standardabweichung gelingt die Überführung des ebenfalls standardisierten Kriteriummeßwerts $C = S_{z(d)}$ in einen Korrelationskoeffizienten zwischen den zeitverschobenen Beobachtungsreihen. Eine rechnerische Bestimmung der Änderungssensitivität eines Items läßt sich demnach über das Kriterium $C = 2(1 - r_{z,z'})$ vornehmen“ (PETERMANN und KNOPF 1976, S. 3).

Beispiel:

Dieses Konzept der änderungssensitiven Items wurde von MEHNERT (nach BASTINE 1970, S. 537) im Bereich der klientenzentrierten Gesprächspsychotherapie erprobt. Ihr Veränderungsfragebogen erfaßte ausschließlich Items, die bereits als Veränderungsaussagen formuliert waren. So trat etwa an die Stelle einer Feststellung wie „Ich bin heiter“, die vor und nach der Behandlung zu beantworten war, das Item „Ich bin heiterer geworden“. Diese Veränderungsaussagen wurden nach der Therapie beantwortet, wobei die Patienten das Ausmaß ihrer Veränderungen über eine mehrstufige Skala beurteilten.

Zweifellos erhält man durch diese Erweiterung zuverlässigere Messungen der klinischen Prozesse. Fraglich bleibt, ob der Test noch den herkömmlichen Validitätskriterien genügt. Trotz dieser Problematik sollte die Änderungssensitivität als zusätzliches Kriterium bei der Konstruktion von „Veränderungstests“ in Zukunft verstärkt diskutiert werden.

7. Reliable Erfassung von Veränderung in der klinischen Forschungspraxis

An dieser Stelle soll zur Ausgangsfrage der Darstellung zurückgekehrt werden:

„WIE KOMME ICH ZU MESSWERTEN, DIE EINE STATTGEFUNDENE VERÄNDERUNG WIDERSPIEGELN, UND WIE KANN MAN SIE INTERPRETIEREN?“

Neben der Diskussion der Veränderungsmaße (vgl. besonders die Erweiterung von CRONBACH und FURBY 1970) ist für die Beantwortung der Frage auch Tabelle

3 von Nutzen. Sie zeigt, daß eine grundsätzliche Entscheidung zwischen Einzelfallbetrachtung und gruppenstatistischem Vorgehen getroffen werden muß. Während in der klinischen Forschungspraxis die gruppenstatistischen Verfahren bisher bevorzugt wurden, muß dennoch DUKES (1965, abgedruckt in diesem Band) zugestimmt werden, der die Brauchbarkeit der Einzelfallbetrachtung für die klinische Forschung ausdrücklich betont. Man sollte sich dabei nicht von den damit verbundenen Problemen (z. B. Stichprobenprobleme) abschrecken lassen (vgl. hierzu DAHME, LEISTIKOW, CATTELL sowie HOLTZMAN, alle in diesem Band).

Nach dieser Entscheidung für oder gegen gruppenstatistische Verfahren sollte man sich kritisch mit der Wahl einer Auswertungsstrategie auseinandersetzen. Der oft beobachtete Trend zu sehr komplexen Verfahren (vgl. etwa kanonische Korrelationsanalyse oder Faktorenanalyse) ist meist der vorliegenden Datenqualität nicht angemessen. Es soll daraus jedoch nicht geschlossen werden, daß man sich bei klinischen Fragestellungen von vornherein mit der Anwendung relativ voraussetzungsloser nonparametrischer Tests begnügen sollte. Ein solcher Standpunkt würde jede weitergehende methodische Entwicklung einschränken. Zu empfehlen sind weniger komplexe Verfahren, die dennoch Kausalaussagen zulassen, wie beispielsweise die Cross-Lagged-Analysis. Dieses Modell ist aufgrund seiner Spezifität in der Lage, (allerdings relativ einfache) Veränderungsaussagen zu überprüfen.

Resümee

Für die klinische Forschungspraxis wird in methodischer Hinsicht eine Entwicklung in verschiedenen Richtungen angestrebt werden müssen, um zu reliablen Veränderungsaussagen zu kommen, wie beispielsweise:

- a) die Präzisierung und Überprüfung der klassischen experimentellen Techniken (vgl. BERNSTEIN und PAUL 1971, NAMBOODIRI in diesem Band, KIRCHNER, KISSEL, PETERMANN und BÖTTGER 1976),
- b) die Erprobung und der verstärkte Einsatz von Techniken der Einzelfallbetrachtung (vgl. DAVIDSON und COSTELLO 1969),
- c) der Einsatz von gruppenstatistischen Verfahren, die von ihren Annahmen über die Datenqualität her klinischen Fragestellungen gerecht werden,
- d) die Erweiterung der klassischen Testtheorie; dies einmal in ihrem eigenen Rahmen, etwa durch die Konstruktion von Veränderungstests (vgl. BEREITER 1963, RENN 1973), zum anderen durch Ansätze, denen Annahmen der probabilistischen Testtheorie (vgl. FISCHER 1974, FRICKE 1972, 1974, HEHL und PETERMANN in diesem Band) zugrundeliegen.