

**Früherkennung der Schizophrenie anhand
subjektiver Beschwerdeschilderungen:
ein methodenkritischer Vergleich der Vor-
hersageleistung nonparametrischer stati-
stischer und alternativer Verfahren zur
Generierung von Vorhersagemodellen.**

von

Frauke Schultze-Lutter, Dipl.-Psych.

geb. 08.02.1966

zur Promotion an der

Philosophischen Fakultät der Universität zu Köln

betreut von

Prof. Dr. G. Fischer

Institut für Klinische Psychologie & Psychotherapie

„So läuft diese Wissenschaft, die mich alles lehren sollte, schließlich auf
eine Hypothese hinaus, die Klarheit taucht in einer
Metapher unter, die Ungewißheit stellt sich als ein Kunstwerk heraus. Hatte
ich so viele Anstrengungen nötig?“

(Albert Camus, 1942)

INHALTSVERZEICHNIS

1	EINFÜHRUNG	1
2	THEORETISCHER UND EMPIRISCHER HINTERGRUND	4
2.1	Früherkennung der Schizophrenie	4
2.1.1	Historische Grundlagen	4
2.1.2	Die Vulnerabilitätshypothese	6
2.1.3	Epidemiologie und Verlauf	9
2.1.4	Präventionsansätze	13
2.1.5	Das Früherkennungsparadigma	15
2.1.5.1	Vulnerabilitätsindikatoren	16
2.1.5.2	Prodromalsymptome	18
2.1.5.3	Methoden zur Erfassung von Prodromalsymptomen	21
2.1.5.4	Entwicklungsmodelle psychotischer Erstmanifestationen	23
2.1.6	DSM-III-R-Prodromalsymptome einer schizophrenen Störung	28
2.1.7	Das Basissymptom-Konzept schizophrener Erkrankungen	31
2.1.7.1	Entwicklungsmodell schizophrener Psychosen nach dem Basissymptom-Konzept	33
2.1.7.2	Schizophreniespezifität von Basissymptomen	36
2.1.7.3	Basissymptome als Prodromalsymptome schizophrener Erkrankungen	44
2.2	Mathematische Methoden zur Generierung von Vorhersagemodellen	48
2.2.1	Logistische Regressionsanalyse	49
2.2.2	Nonparametrische Diskriminanzanalyse	53
2.2.3	Rekursive Partitionierung	55
2.2.3.1	Search Partition Analysis	61
2.2.4	Grob-Mengen-Theorie	65
2.2.4.1	Grundlagen	66
2.2.4.2	DataLogic	71
2.2.5	Diagnostische Effizienzindizes	77
2.2.5.1	Diagnostische Likelihood Ratios	81
2.2.6	Vergleichsstudien	83
2.3	Fragestellung	94
3	METHODE	95
3.1	Versuchsplanung (1)	95
3.1.1	Versuchsplan-Anlage (1)	95

3.1.2	Daten	95
3.1.2.1	Stichprobe	96
3.1.2.2	Instrumente	99
3.1.2.3	Untersuchungsdurchführung	103
3.1.2.4	Eignung der Stichprobe	103
3.1.2.5	Stichprobenteilung	104
3.2	Versuchsplanung (2)	107
3.2.1	Bedingungsvariationen der unabhängigen Variable 'Methode'	107
3.2.2	Versuchsplan-Anlage (2)	110
3.2.3	Ableitung von empirischen Vorhersagen	112
3.2.4	Ableitung von statistischen Vorhersagen	113
3.2.5	Ableitung von statistischen Hypothesen	114
3.2.6	Testplanung	114
4	ERGEBNISSE	115
4.1	Ergebnisse der vorgeschalteten Analysen	115
4.1.1	Häufigkeitsvergleich	115
4.1.2	Clusteranalyse	115
4.2	Vorhersagemodelle	119
4.2.1	Diskriminanzanalyse der Einzelitems	119
4.2.2	Diskriminanzanalyse der Cluster	122
4.2.3	Logistische Regressionsanalyse der Einzelitems	124
4.2.4	Logistische Regressionsanalyse der Cluster	126
4.2.5	Grob-Mengen-Modell	127
4.2.6	Search Partition Analysis	129
4.2.7	Diagnostische Effizienzindizes	131
4.2.8	Diagnostische Likelihood Ratios	133
4.2.9	Vergleich der ausgewählten Variablen	135
4.3	Klassifizierungsleistung der Vorhersagemodelle	135
4.4	Plausibilität der Vorhersagemodelle	140
4.5	Anwendbarkeit der Vorhersagemodelle	143
5	DISKUSSION	145
6	ZUSAMMENFASSUNG	167
7	LITERATUR	170

ANHANG		189
A	Bonner Skala für die Beurteilung von Basissymptomen, BSABS	189
B	Aachener Kurzversion der BSABS	198
C	72-Item Kurzversion der BSABS	201
D	Beurteilungsbogen für Entscheidungsregeln	204
E	Häufigkeitsvergleich aller Einzelvariablen in den beiden Outcome-Gruppen	207
F	Hierarchische Clusteranalyse: Dendogramm der Fusionierungsstufen	211
G	Entwicklung der Fehlerquadratsumme: 'Elbow'-Kriterium zur Bestimmung der Clusteranzahl	212
H	Korrelationsmatrix zwischen Einzelvariablen und Clustern	213
I	Diagnostische Gütekriterien der Einzelvariablen	216
J	Übersicht über die in die einzelnen Vorhersagemodelle aufgenommenen Variablen	219
K	Übersicht über die Gruppenzuordnungen der einzelnen Vorhersagemodelle für die Probanden der Validierungsstichprobe	220
L	Lebenslauf	222

1 EINFÜHRUNG

In der Medizin gilt der Prophylaxe, der Frühdiagnose und der Prävention schwerwiegender oder häufig vorkommender Erkrankungen, beispielsweise Karies, Parodontitis marginalis, verschiedensten Malignomen oder Demenzen vom Alzheimer-Typ, bereits seit langem ein besonderes Augenmerk. Diesen Erkrankungen ist gemeinsam, daß sie von einem bestimmten Ausprägungsgrad an kaum mehr abzufangen und gravierende, kostenträchtige Folgeschäden oft unvermeidbar sind. Obwohl dies auch für eine Reihe von psychischen Störungen gilt, und insbesondere für schizophrene Psychosen, wurde in der Psychiatrie lange Zeit der Präventivforschung wenig Beachtung geschenkt. Die Ursache hierfür lag darin, daß bei der ohnehin in diesem Gebiet oft schwierigen Differentialdiagnostik eine Früherkennung schwerer psychischer Erkrankungen, wie sie etwa die schizophrenen Psychosen darstellen, in den eher uncharakteristischen Prodromalphasen von der großen Mehrheit der Forscher als aussichtslos angesehen wurde.

In den letzten Jahren hat sich diese Einstellung insbesondere durch Erfolge auf dem Gebiet der Früherkennung drohender Rezidive und deren oft erfolgreicher Prävention drastisch verändert: Eine frühzeitige Erkennung und Behandlung schizophrener Psychosen vor der psychotischen Erstmanifestation wird heute als ein realistisches Forschungsziel angesehen und findet weltweit steigendes Interesse. In der Früherkennungs- und -interventions-Forschung schizophrener Psychosen ergeben sich jedoch nicht nur eine Reihe konzeptioneller Fragen - zum Beispiel die Frage nach der Bedeutung eines (potentiellen) Prodromalsymptoms für eine Psychose und seine Beziehung zum Risiko einer nachfolgenden psychotischen Dekompensation - und praktischer Unwägbarkeiten - etwa die Identifizierung von Personen in einer Prodromalphase und ihre Einbindung in sekundärpräventive Maßnahmen, sondern auch methodisch-mathematische Probleme.

Dieser letzte Punkt bezieht sich neben Fragen der Versuchsplanung vor allem auf die Entwicklung und Überprüfung von Techniken und Strategien zur Bestimmung eines bestmöglichen und ökonomischen Vorhersagemodells für eine dichotome Zielvariable - zum Beispiel 'spätere Entwicklung einer schizophrenen Psychose' versus 'keine spätere Entwicklung einer schizophrenen Psychose' - aus einer größeren Anzahl potentiell prädiktiver Variablen. Während bei Entwicklung solcher Verfahren in den letzten zwei Jahrzehnten große Fortschritte erzielt wurden, haben sich bisher nur wenige Studien dem Vergleich der unterschiedlichen Techniken hinsichtlich ihrer Vorhersageleistung für einen alternativen Ausgang (engl.: „outcome“) gewidmet.

Um die praktische Relevanz solcher Vergleiche zu erfassen, ist es sicherlich notwendig, sich vor Augen zu führen, daß das Ergebnis der verwandten Technik, d.h. das resultierende Vor-

hersagemodell, die spätere Beurteilung von Einzelpersonen maßgeblich bestimmen wird. Hierbei können schon geringfügige Unterschiede in der Vorhersageleistung einzelner Methoden zur Generierung von prognostischen Modellen im Einzelfall darüber entscheiden, ob bei einem Patienten eine beginnende Schizophrenie - oder andere Erkrankung - zu Recht oder Unrecht (nicht) diagnostiziert wird. Da diese diagnostische Einschätzung die Grundlage der weiteren Behandlung bildet, sind die Folgen einer Fehlbeurteilung im Einzelfall erheblich:

So kann etwa mit der fälschlichen Diagnose eines schizophrenen Prodroms eine nicht indizierte Medikation mit Neuroleptika einher gehen, die mit dem Risiko potentieller Nebenwirkungen und Spätfolgen verbunden ist. Darüber hinaus wird der betroffene Patient möglicherweise einer unnötigen Stigmatisierung ausgesetzt, die auch heute noch die Diagnose einer Schizophrenie oft mit sich bringt, und die Einleitung einer adäquaten Behandlung der eigentlichen Erkrankung bzw. die Fokussierung auf den eigentlichen Problembereich verzögert.

Zu einer solchen Verzögerung der Einleitung einer adäquaten Behandlung führt auch die Verkennung eines schizophrenen Prodroms. Die Folgen einer solchen Verzögerung sind eine möglicherweise verzögerte und unvollständige Remission der psychotischen Symptomatik, eine längere stationäre Behandlungsbedürftigkeit, ein höheres Rückfallrisiko, ein steigendes Risiko der Entwicklung persistierender Negativsymptomatik, eine größere Belastung des Betroffenen, seines Umfelds und seiner Arbeits- und Ausbildungssituation, ein erhöhtes Risiko zur Ausbildung sozialer Defizite, depressiver Symptomatik und suizidaler Tendenzen, eine Steigerung der Neigung zu Substanzmißbrauch und nicht zuletzt deutlich höhere Behandlungskosten.

In der vorliegenden Arbeit soll deshalb anhand der Daten einer prospektiven psychometrischen Studie, des Köln-Bonn-Aachener Früherkennungs-Projekts, die Vorhersage- bzw. Klassifizierungsleistung von acht Strategien zur Variablenselektion und Entwicklung multivariater Vorhersagemodelle verglichen werden, die zwischen Übergang und Nicht-Übergang in eine Schizophrenie unterscheiden. Bei den untersuchten Verfahren handelt es sich um die inferenzstatistischen Ansätze der logistischen Regression und der nonparametrischen Diskriminanzanalyse - jeweils mit und ohne vorgeschaltete hierarchische Clusteranalyse, um zwei eher deskriptive Ansätze, die zum einen auf den 'klassischen' diagnostischen Effizienzindizes und zum anderen auf den alternativ hierzu vorgeschlagenen diagnostischen Likelihood Ratios basieren, um einen mengentheoretisch basierten Ansatz nach der Grob-Mengen-Theorie von Pawlak (1984) und um die auf binäre Daten spezialisierte nicht-hierarchische Partitionierungsanalyse von Marshall (1986, 1995), die Elemente der Inferenzstatistik mit Booleschen und mengentheoretischen Überlegungen verbindet. Die beiden letztgenannten Verfahren haben bereits in der Inneren und Zahnmedizin sowie der Onkologie

erfolgreich Anwendung gefunden, wurden jedoch in der psychiatrischen Forschung bislang nicht berücksichtigt. Ihrer Beschreibung wird deshalb besondere Aufmerksamkeit gewidmet, während die bekannteren inferenzstatistischen Verfahren nur kurz skizziert werden.

Vor der Darstellung der verwandten Methoden in Kapitel 2.2 wird einleitend ein Überblick über die Grundlagen und den heutigen Stand der Früherkennung in der Schizophrenieforschung gegeben und das den Daten zugrundeliegende Basissymptomkonzept erläutert.

Kapitel 3 wird sowohl den Aufbau der Studie als auch die Versuchsplanung und -durchführung behandeln. Daran anschließend werden in Kapitel 4 die Ergebnisse der einzelnen Evaluationsverfahren dargestellt und im Hinblick auf ihre Vorhersagekraft für den Übergang in eine schizophrene Psychose sowie auf ihre inhaltliche Plausibilität und Anwendbarkeit verglichen. Die Ergebnisse und die daraus resultierenden Implikationen für die Auswahl zukünftiger Evaluationsmethoden in diesem Forschungsbereich werden in Kapitel 5 diskutiert. Eine abschließende Zusammenfassung der Arbeit wird in Kapitel 6 gegeben.

2 THEORETISCHER UND EMPIRISCHER HINTERGRUND

2.1 FRÜHERKENNUNG SCHIZOPHRENER PSYCHOSEN

2.1.1 Historische Grundlagen

Der Begriff *Schizophrenie* (griech.: „schizo“ = ich spalte, „phren“ = Geist) wurde 1908 von dem Schweizer Psychiater Eugen Bleuler (1857-1939) eingeführt. Er bezeichnete hiermit diejenigen psychischen Störungen, deren Zusammenfassung unter der Bezeichnung *Dementia praecox* zwölf Jahre zuvor von Emil Kraepelin (1896; 1854-1925) vorgelegt worden war, schlug aber den Terminus Schizophrenie mit der Begründung vor, daß die Spaltung psychischer Funktionen ein kennzeichnendes Symptom dieser Diagnosegruppe sei.

Kraepelin hatte eine Unterscheidung der heute als *funktionelle* oder *endogene Psychosen* bezeichneten Krankheitsbilder in unabhängige, nach Symptomatik, Ätiologie, die von ihm bereits auf physiologischer Ebene angenommen wurde, und klinischem Verlauf differenzierbare Krankheitsentitäten angestrebt und war zu einer Dichotomisierung in affektive Störungen (Melancholie und Manie) und *Dementia praecox* (Katatonie, Hebephrenie und Paranoia) gelangt. Obwohl ihm bewußt war, daß eine kleine Gruppe von Patienten, die zunächst manifeste Symptome einer *Dementia praecox* aufgewiesen hatten, keinen geistigen Abbau zeigten, postulierte er als Charakteristikum dieses Krankheitsbildes einen progressiven, chronischen Krankheitsverlauf, der unausweichlich in eine 'Verblödung', eine Demenz führe.

Kraepelins Konzept und seine pessimistische Sicht des Verlaufs beeinflusste die traditionelle Psychiatrie lange Zeit maßgeblich. Auch heute, in der im Zuge der Entwicklung neuerer biologischer Ansätze in der Ätiologieforschung entstandenen Neo-Kraepelinschen Ära, sind die darauf basierenden Symptomanordnungen von Kurt Schneider (1887-1967) bestimmend geblieben. In enger Anlehnung an Kraepelins Symptomaufzählung und dessen Kriterium eines chronischen Verlaufs erarbeitete Schneider (1946) eine Klassifizierung der schizophrenen Symptomatik in Symptome ersten und zweiten Ranges, womit er gleichzeitig eine diagnostisch relevante Gewichtung einführte. Unter die Symptome ersten Ranges faßte er Gedankenlautwerden, -entzug, -ausbreitung und -eingebung, dialogische, imperative und kommentierende Stimmen, leibliche Beeinflussungserlebnisse, Willensbeeinflussung und Wahrnehmung; zu den Symptomen zweiten Ranges zählte er optische, olfaktorische, gustatorische und sonstige akustische Halluzinationen, Wahneinfall, Coenästhesien im engeren Sinne, einfache Eigenbeziehung, Ratlosigkeit, depressive und maniforme Verstimmungen sowie eine subjektiv erlebte Gefühlsverarmung.

Auch E. Bleuler hatte in seiner Monographie von 1911 schon eine diagnostische Gewichtung schizophrener Symptome vorgenommen, indem er zwischen (primären) Grundsymptomen

und (sekundären) akzessorischen Symptomen differenzierte. Seine Ansicht, es handele sich bei Schizophrenien nicht um einen tatsächlichen Intelligenzverfall, eine Demenz, sondern um eine Minderung der Zielgerichtetheit von Denkvorgängen bei überwiegend intakten grundlegenden Sinnesfunktionen bestimmte seine Definition von Assoziationsstörungen, Affektstörungen, intellektueller Ambivalenz und Ambitendenz sowie Autismus als Grundsymptome. Alle anderen Symptome, insbesondere Halluzinationen, Wahn und katatone Symptome, betrachtete er als sekundär. Obwohl er weitgehend Kraepelins pessimistische Verlaufsbeurteilung teilte, schloß Bleuler im Gegensatz zu Kraepelin Patienten mit einer guten Prognose nicht aus und erweiterte die Gruppe der Schizophrenien um Patienten mit atypischen Manien und Melancholien, psychotischer Symptomatik und auffälliger nervöser, zwanghafter oder impulsiver Symptomatik. Den drei Subtypen der *Dementia praecox* fügte er darüber hinaus die *Schizophrenia simplex* hinzu, wie sie auch heute noch in den diagnostischen Leitlinien der Weltgesundheitsorganisation (WHO), der 'International Classification of Diseases - ICD-10' (1993), unter F20.6 aufgeführt ist, und prägte schließlich den Begriff der *latenten Schizophrenie*.

Bleuler definierte die latente Schizophrenie als feinen, oft subklinischen psychopathologischen Ausdruck der Disposition zur manifesten Schizophrenie, wie er als *formes frustes* gerade im familiären Umfeld Schizophrener und im individuellen Vorfeld der psychotischen Dekompensation gehäuft vorkomme. Diese Konzeption hat im ICD-10 unter der Bezeichnung *schizotype Störung* (Kodierung F21) Eingang gefunden, wo sie der Schizophrenie und den wahnhaften Störungen zur Seite gestellt wurde.

Die Bezeichnung *schizotype Störung* knüpft an Meehls (1962, 1990) in der anglo-amerikanischen Psychiatrie höchst einflußreiche Konzeption der *Schizotypie* an. Schizotypie meint hierbei eine Persönlichkeitsdisposition für die Entwicklung einer Schizophrenie, die auf der Grundlage eines hypothetischen genetisch fundierten neurologischen Zustandes, der *Schizotaxie*, durch den Einfluß polygenetisch determinierter kognitiver Potentatoren und ungünstiger psychologischer Interaktionen mit der Umwelt entsteht. Der schizotypen Störung nach ICD-10 entspricht in den diagnostischen Leitlinien der American Psychiatric Association (APA), der vierten Fassung des 'Diagnostic and Statistical Manual of Mental Disorders - DSM-IV' (1994) weitgehend die *schizotypische Persönlichkeitsstörung* (301.22). Anders als in der ICD wurde die schizotype Persönlichkeitsstörung im DSM hingegen in Anlehnung an Meehls Einschätzung der Schizotypie als Persönlichkeitsdisposition diagnostisch den Persönlichkeitsstörungen zugeordnet. Von der Sache her ist mit Meehls Schizotypie und Bleulers latenter Schizophrenie jedoch etwas ganz Ähnliches gemeint (Klosterkötter, 1998).

2.1.2 Die Vulnerabilitätshypothese

Gegenüber dem Lebenszeitrisko für eine schizophrene Erkrankung in der Allgemeinbevölkerung von einem Prozent zeigten epidemiologische Studien ein erhöhtes Erkrankungsrisiko bei Personen mit einer familiären Belastung. In einer Übersicht von Familien- und Zwillingsstudien, die in Europa zwischen 1920 und 1987 durchgeführt wurden, fand Gottesman (1991) ein durchschnittliches Lebenszeitrisko für die Entwicklung einer schizophrenen Psychose bei biologisch Verwandten bereits manifest schizophren erkrankter Personen, das in Abhängigkeit vom Verwandtschaftsgrad zwischen 2% bei Cousins/Cousinen dritten Grades und 48% bei monozygoten Zwillingen liegt.

Die Tatsache, daß die Genpenetranz auch bei monozygoten Zwillingen weit unter 100% liegt, leistete dem Entstehen von Vulnerabilitätsmodellen Vorschub, wie etwa dem kausalen Modell von Meehl (1962). Den verschiedenen Vulnerabilitätsmodellen ist die Vorstellung gemeinsam, daß die Anwesenheit einer genetisch determinierten Anfälligkeit allein nicht ausreichend ist, um den Genotyp in einen Phänotyp umzusetzen, so daß die Einführung eines auslösenden Moments erforderlich ist, das in Form eines Stressors eine bereits vorhandene Vulnerabilität als sichtbare Störung manifestiert (s. Abb. 1 und 2).

Eine der frühesten und heute immer noch zutreffenden Überlegungen zur Vulnerabilitätshypothese geht auf Karl Friedrich C. Canstatt (1807-1850) und seine Beschäftigung mit den Ursachen der „Geisteskrankheiten“, die heute weitgehend unter die Psychosen subsumiert würden, zurück:

Man beobachtet bei vielen Individuen, die man noch nicht geradezu geisteskrank nennen kann, einen solchen Hochstand der psychischen Erregung, daß es nur eines geringen Anlasses bedarf, damit wirkliche Alienation entstehe. Wir nennen dies die psychische Vulnerabilität und die meisten Gelegenheitsursachen, welche die Geisteskrankheit ins Dasein rufen, finden bereits diese Prädisposition vor. (Canstatt, 1841, S.329)

Im Gegensatz zu Kraepelins Krankheitsmodell der Dementia praecox setzen Vulnerabilitätsmodelle nicht die Existenz eines zugrunde liegenden, unvermeidlichen und progressiven Krankheitsprozesses voraus. Sie unterscheiden jedoch den Schizophrenen von anderen Personen, indem sie bei ihm eine überdauernde Prädisposition postulieren, die unter bestimmten psychosozialen Bedingungen die Entwicklung einer akuten Krankheitsepisode begünstigt. Hinweise auf eine genetische Belastung finden sich indes bei der Mehrzahl von schizophren Erkrankten nicht. Bei solchen "sporadisch Schizophrenen" oder "Phänokopien" (Zubin, 1990, S.46) wird eine erworbene Vulnerabilität vermutet, die z.B. durch pränatale oder perinatale Läsionen hervorgerufen sein kann (Murray, 1994).

Das heute weltweit anerkannteste und die Ergebnisse der Schizophrenieforschung wohl am besten integrierende Vulnerabilitätsmodell stammt in seiner heutigen Fassung von Nuechterlein und Dawson (1984; Nuechterlein, 1987; Nuechterlein & Zaucha, 1990). Es postuliert, daß das Zusammenspiel vielfältiger Risiko- und Schutzfaktoren - der Person inhärente Vulnerabilitätsfaktoren, umweltbedingte Potentatoren und Stressoren sowie personbezogene und umweltbedingte Schutzfaktoren (s. Abb. 1) - in einer bestimmten Lebenssituation darüber entscheidet, ob sich über sogenannte *intermediäre 'states'* (Zwischenstufen) Prodromalsymptome und daraus schließlich die akute psychotische Episode mit ihren möglichen Begleitsymptomen entwickelt.

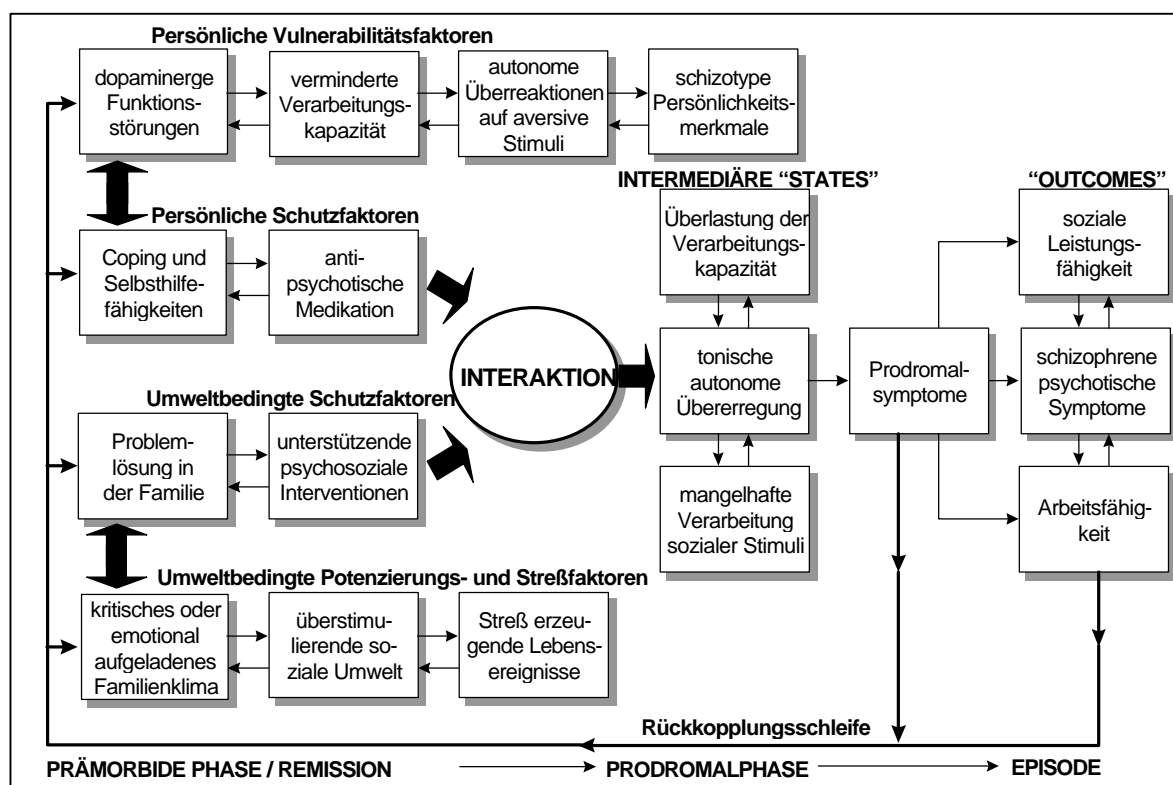


Abb. 1: Das Vulnerabilitäts-Streß-Bewältigungs-Modell für die Entwicklung schizophrener Episoden (nach Nuechterlein, 1987)

Entstehen schizophrene Episoden ohne einen erkennbaren äußeren Anlaß, d.h. spielen persönliche Vulnerabilitätsfaktoren eine entscheidende Rolle, so ist nach dem Vulnerabilitäts-Streß-Bewältigungs-Modell von einer hohen genetischen und/oder erworbenen Vulnerabilität auszugehen. Sind jedoch umweltbedingte Potenzierungs- und Stressoren im Vorfeld einer schizophrenen Episode wirksam, wird umgekehrt eine geringe genetisch-hirnorganische Vulnerabilität unterstellt, die durch psychosoziale Faktoren erst über die Schwelle zur schizophrenen Symptombildung angehoben wird. Die kritische Schwelle ist hierbei generell um so

höher anzusetzen, je stärker persönliche und umweltbedingte Schutzfaktoren ausgebildet sind (Klosterkötter, 1996). Das Zusammenspiel dieser drei Hauptfaktoren, Vulnerabilitäts-, Schutz- und Potenzierungs- bzw. Streßfaktoren, wird in Anlehnung an Leffs Schwellenmodell (Leff, 1987) in Abbildung 2 verdeutlicht.

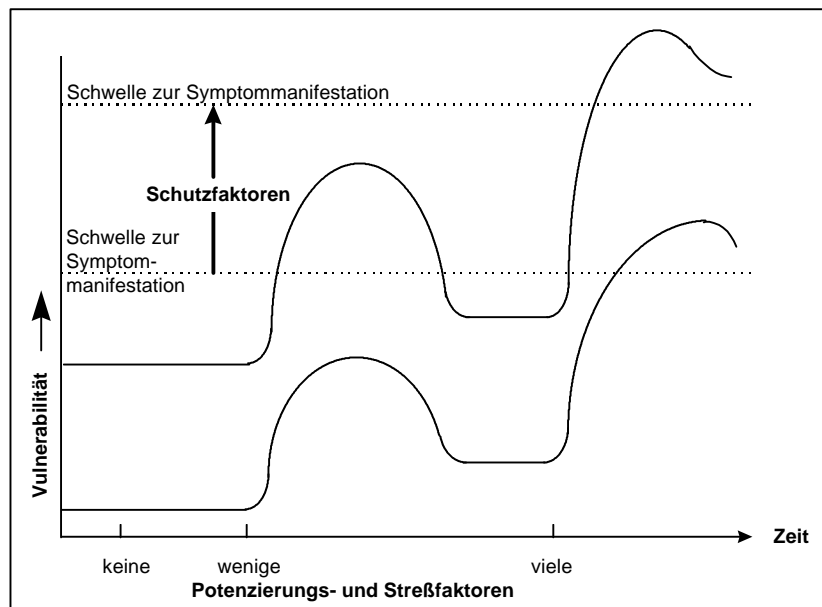


Abb. 2: Schwellenmodell der Interaktion zwischen Vulnerabilitäts-, Schutz- und Potenzierungs- bzw. Streßfaktoren anhand eines hypothetischen Verlaufs mit geringer Vulnerabilität (untere Kurve) und eines mit hoher Vulnerabilität (obere Kurve; in Anlehnung an Leff, 1987).

Ob eine Person anfällig für eine schizophrene Psychose ist, hängt also entscheidend von den Vulnerabilitätsfaktoren ab. Während sich die ersten drei persönlichen Vulnerabilitätsfaktoren (s. Abb. 1) auf psychophysiologisch meßbare Auffälligkeiten beziehen, wird mit dem vierten Faktor, den schizotypen Persönlichkeitsmerkmalen, die psychometrische Untersuchungsebene angesprochen. Die Gleichstellung dieses hypothetischen Vulnerabilitätsfaktors mit schizotypen Persönlichkeitsmerkmalen begründet sich in der Überlappung der bei 'high-risk'-Kindern mit mindestens einem schizophrenen Elternteil gefundenen affektiv-kognitiven Störungsmuster aus emotionaler Labilität, sozialer Ängstlichkeit, Aggressivität und Mißtrauen, verflachtem Affekt, eigenartigem Verhalten, Wahrnehmungsabweichungen, bizarren Denkinhalten und sozialer Anhedonie (Olin & Mednick, 1996; Parnas, 1999) mit den DSM-IV-Kriterien einer schizotypischen Persönlichkeitsstörung. So ist aus der Sicht der 'high-risk'-Forschung ohnehin die Schizotypie als Schizophrenie-Spektrum-Störung aufzufassen, wobei angenommen wird, daß sowohl Personen mit Schizophrenie als auch solche mit schizotypischer Persönlichkeitsstörung, sogenannte "*schizotypes*" (Olin et al., 1997, p. 93), eine ver-

gleichbare genetische Anfälligkeit für eine schizophrene Erkrankung besäßen. Die Interaktion mit der Umwelt stellt nach dieser Konzeption die entscheidende Determinante dar, die darüber entscheidet, ob ein solcher Schizotyp im Laufe seines Lebens psychotisch dekompenziert.

Entsprechend rechnet Nuechterlein zu den schizotypen Persönlichkeitsmerkmalen auch subklinische kognitive Gleitvorgänge, wie sie etwa mit dem 'Thought Disorder Index - TDI' (Johnston & Holzman, 1979) erfaßt werden, und feine selbstwahrgenommene Denk-, Perzeptions- und Propriozeptionsstörungen im Sinne der sogenannten *Wisconsin-Skalen*, insbesondere der 'Physical Anhedonia Scale - PAN' (Chapman, Chapman & Raulin, 1976), der 'Magical Ideation Scale - MAG' (Eckblad & Chapman, 1983) und der 'Perceptual Aberration Scale - PER' (Chapman, Chapman & Raulin, 1978), die stark auf Meehls Konzeption der Schizotypie basieren (eine neuere Übersicht über die einzelnen Skalen und hierauf bezogene Forschungsergebnisse gibt Edell, 1995). Solche selbstwahrgenommenen subklinischen attentional-perzeptiv-kognitiven Störungen sind unter dem Terminus technicus *Basissymptome* bereits seit den 60er Jahren auch in Deutschland von Gerd Huber (1966) beschrieben und später in Form der 'Bonner Skala für die Beurteilung von Basissymptomen - BSABS' (Gross, Huber, Klosterkötter & Linz, 1987) operationalisiert worden (vgl. Kap. 2.1.7 und 3.1.2.2).

2.1.3 Epidemiologie und Verlauf

In einer vergleichenden Studie der WHO zu Manifestationsformen, Inzidenz und Verlauf von schizophrenen Störungen in 10 Ländern an insgesamt 1 379 schizophrenen Patienten im Alter zwischen 15 und 54 Jahren und einem 2-Jahres-Follow-up (Jablensky et al., 1992) wurden die Ergebnisse der in den 60er Jahren durchgeführten International Pilot Study of Schizophrenia (IPSS; WHO, 1973) weitgehend bestätigt: Schizophrene Erkrankungen traten weltweit mit ähnlichen Inzidenzraten auf (1,5 bis 4,2 von 100 000 Personen bei einer ICD-9 Schizophrenie-Definition (1975), 0,7 bis 1,4 pro 100 000 bei einer engeren Definition nach CATEGO S+ (Wing, 1976)) und beeindruckten in den verschiedenen Kulturen eher durch die Ähnlichkeiten als durch die Unterschiede im klinischen Bild. Das aus den Daten erschlossene Lebenszeitrisko, an einer Schizophrenie zu erkranken, betrug hiernach bei Zugrundelegen der weiter gefaßten ICD-9-Definition zwischen 0,5 und 1,72% und bei Verwendung der strikteren Definition nach CATEGO S+ zwischen 0,26 und 0,54%. Das Alter für Ersterkrankungen lag im Durchschnitt bei Frauen einige Jahre höher als bei Männern, wobei 58% der Patienten jünger als 35 Jahre waren. Der Krankheitsausgang zeigte sich insgesamt variabel

und abhängig von genetischen, entwicklungsmäßigen und umweltbedingten Einflüssen, erschien aber in entwickelten Ländern günstiger. In dem zweijährigen Katamnesezeitraum blieb es bei 50,3% der Patienten bei einer einmaligen psychotischen Episode, 31,1% hatten zwei oder mehr Episoden, die jeweils vollständig remittierten, und nur bei 15,7% kam es zu einem chronischen Verlauf. Zehn bis 19% der Patienten wurden retrospektiv als extrem schüchterne, vornehmlich allein spielende und überangepaßte Kinder beschrieben, und in der Adoleszenz waren retrospektiv 30% der schizophrenen Patienten als sehr reserviert und ruhig, 22% als überaus sensibel und 10% als besonders mißtrauisch in ihrem Verhalten geschildert worden.

Eine in Deutschland mit dem 'Instrument for the Retrospective Assessment of the Onset of Schizophrenia - IRAOS' (Häfner et al., 1990) durchgeführte, methodisch sehr elaborierte epidemiologische Studie, die vornehmlich auf die retrospektive Erfassung des frühen Verlaufs bei Patienten mit einer schizophrenen Erstmanifestation abzielte, die Mannheimer A(ge)-B(eginning)-C(ourse) Study (Häfner, 1993; Häfner et al., 1990; Häfner, Nowotny, Löffler, an der Heiden & Maurer, 1995; Maurer & Häfner, 1995) kam zu ganz ähnlichen Ergebnissen, bot darüber hinaus aber noch weitere Aufschlüsse über den Frühverlauf von Schizophrenien.

Bei 73% der Patienten zeigte sich hiernach eine prodromale Phase von durchschnittlich fünf Jahren (s. Abb. 3), die durch eine kontinuierliche oder häufig wiederkehrende unspezifische Symptomatik und/oder Negativsymptomatik gekennzeichnet war und bei Männern im Schnitt drei bis vier Jahre früher als bei Frauen einsetzte. Die Prodromalphase begann in 77% aller Fälle vor dem 30. Lebensjahr, in 4% sogar vor dem zehnten Lebensjahr. Unter den mit dem IRAOS erhobenen subjektiv erlebten unspezifischen Symptomen befinden sich auch Symptome der BSABS und der Wisconsin-Skalen PER und PAN; die beobachtbaren Negativsymptome wurden u.a. in Anlehnung an die 'Scale for the Assessment of Negative Symptoms - SANS' (Andreasen, 1982) definiert. Hierbei ist zu berücksichtigen, daß zwischen unspezifischen und negativen Symptomen einige Berührungspunkte bestehen und gerade subjektive Defizienzerfahrungen der Patienten einen Teil der schizophrenen Minussymptomatik ausmachen (Kukla & Gold, 1991). In der ABC-Studie fanden sich als häufigste erste Anzeichen der Störung in 22 bis 14% der Fälle Konzentrations- und subjektive Denkstörungen, Antriebsverlust und Verlangsamung, Mißtrauen und sozialer Rückzug sowie Angst im Sinne einer Veränderung gegenüber der prämorbidem Phase (Häfner et al., 1995).

Im Anschluß an die Prodromalphase zeigte sich eine durchschnittlich 1,1 Jahre währende psychotische Vorphase, deren Beginn durch das Auftreten des ersten positiven Symptoms definiert war und in der es zu einem eklatanten Anstieg der Positivsymptomatik kam, so daß

eine stationäre Aufnahme im Schnitt zwei Monaten nach Erreichen der maximalen Ausprägung der Positivsymptomatik erfolgte (s. Abb. 3).

Eine Erhebung des sozialen Funktionsniveaus zwei Jahre nach der Ersthospitalisierung an der gleichen Stichprobe zeigt einen deutlichen Zusammenhang mit dem prämorbidem sozialen Status, wobei Defizite in der sozialen Rollen Anpassung in der Hauptsache während der Prodromal- und der psychotischen Vorphase entstanden. Der im Vergleich zu Frauen eher ungünstigere langfristige Verlauf bei Männern wurde daher auf den früheren Krankheitsbeginn und die dadurch früher und umfangreicher gestörte soziale Anpassung zurückgeführt. Diese Ergebnisse verdeutlichen nach Häfner und Kollegen (1995) die Dringlichkeit der Frage nach einem früheren Einsatz therapeutischer und rehabilitativer Maßnahmen, um der Entstehung und Verfestigung sozialer Defizite entgegen zu wirken.

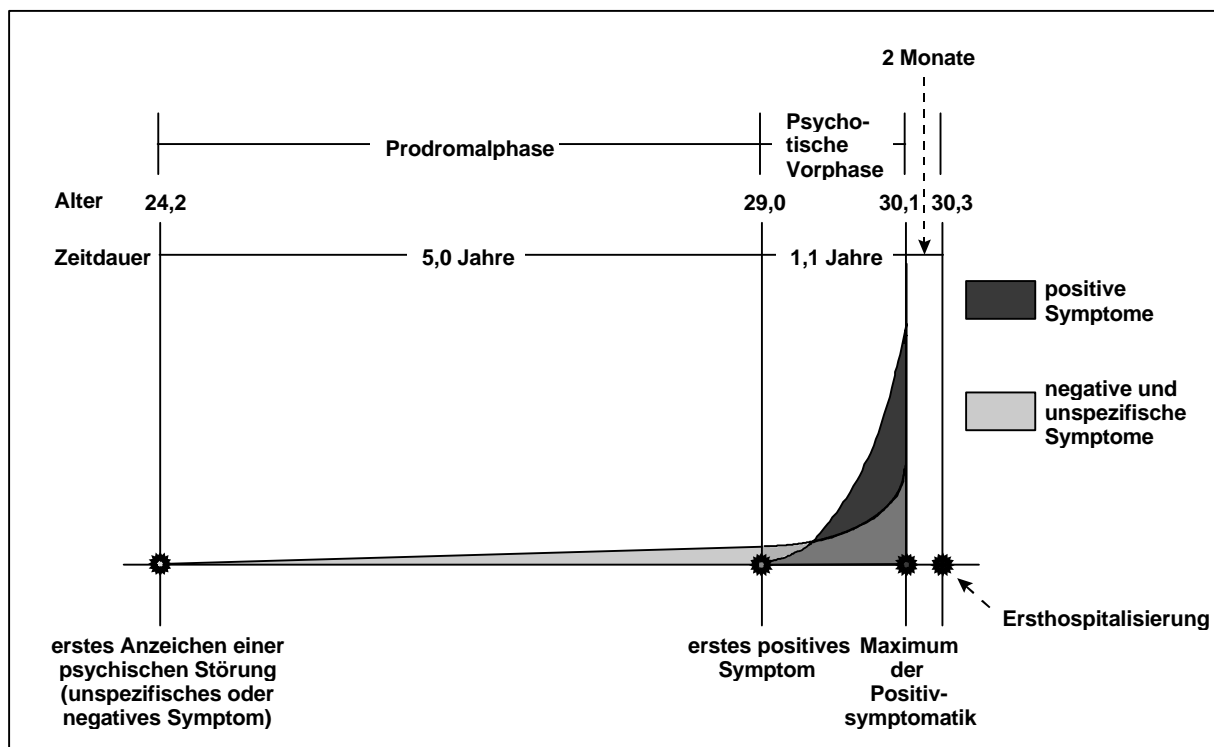


Abb. 3: Frühphasen der Schizophrenie vom ersten Krankheitsanzeichen bis zur Ersthospitalisierung (modifiziert nach Häfner & Maurer, 1996).

Die negativen Auswirkungen des verzögerten Beginns der Behandlung einer schizophrenen Psychose werden ebenfalls durch Ergebnisse der neueren sogenannten *first-episode*-Forschung belegt. Diese Forschungsrichtung zielt insbesondere auf eine Verkürzung der Dauer der psychotischen Vorphase ab, in der bereits schizophrentypische und diagnostisch relevante Positivsymptome vorliegen, da immer wieder gezeigt werden konnte, daß der Zeitraum der unbehandelten oder nicht adäquat behandelten Psychose negativ mit dem

Therapieerfolg korreliert ist (Malla, Norman & Voruganti, 1999). So fanden sich in den unterschiedlichen Arbeiten hierzu Zusammenhänge zwischen der Dauer der unbehandelten Psychose und

- einer verzögerten und unvollständigen Remission der Symptomatik (Birchwood & Macmillan, 1993; Johnstone, Crow, Johnson & Macmillan, 1986; Loebel et al., 1992; McGorry, Edwards & Mihalopoulos, 1996),
- einer längeren stationärer Behandlungsbedürftigkeit und einem höheren Rückfallrisiko (Helgason, 1990),
- einer geringeren Compliance, einer höheren Belastung der Familie und einem erhöhten Expressed Emotion-Niveau (Brown & Birtwistle, 1998; Stirling et al., 1991, 1993),
- einem erhöhten Komorbiditäts- und Suizidrisiko (Addington & Addington, 1998; Addington, Addington, & Patten, 1998; Hambrecht & Häfner, 1996; Koreen et al., 1993; Strakowski, Keck, McElroy, Lonczak & West, 1995),
- einer größeren Belastung der Arbeits- und Ausbildungssituation (Johnstone, Macmillan, Frith, Benn & Crow, 1990; Larsen, McGlashan & Moe, 1996)
- einem schwächeren supportiven sozialen Netzwerk (Larsen, Johannessen & Opjordsmoen, 1998),
- erhöhtem Substanzmißbrauch und delinquentem Verhalten (Humphreys, Johnstone, Macmillan & Taylor, 1992),
- möglichen zerebralen pathophysiologischen Veränderungen (Lieberman, Kinon & Loebel, 1990; Wyatt, 1991) und
- höheren Behandlungs- und Folgekosten (Genduso & Haley, 1997; McGorry & Edwards, 1997; Williams & Dickson, 1995).

Da diese Ergebnissen jedoch in erster Linie auf korrelativen Zusammenhängen basieren, ist zu beachten, daß rein korrelative Beziehungen noch keine Aussagen über kausale Zusammenhänge erlauben und auch durch eine dritte Variable konfundiert sein können, wie etwa die Art des Krankheitsbeginns - schleichend vs. akut (Malla & Norman, 1999). Dennoch kommen Malla, Norman und Voruganti (1999) nach einer Sichtung der first-episode-Literatur zu dem Schluß: „The arguments for early and optimal intervention are compelling.“ (p. 845). Und es erscheint wie bei vielen somatischen Erkrankungen auch bei der Schizophrenie der Grundsatz Gültigkeit zu besitzen: „*Je früher die Behandlung einsetzt, um so günstiger gestaltet sich der Verlauf.*“ (Klosterkötter, 1998, S. 367), wobei der genaue Zeitpunkt, zu dem eine solche Behandlung einsetzen sollte, noch weiterer Forschung bedarf, um etwa eine beginnende akut-psychotische Störung klar von einer kurzen, spontan remittierenden psychotischen Episode abgrenzen zu können (Malla & Norman, 1999).

2.1.4 Präventionsansätze

Den Ergebnissen von Studien zum Verlauf schizophrener Störungen folgend muß hiermit in den meisten Fällen eine schizophrene Erkrankung als ein Prozeß aufgefaßt werden, der unterschiedliche Phasen durchläuft. Entsprechend kann die Prävention eines Ausbruchs einer schizophrenen Episode auf verschiedenen Ebenen erfolgen: der primären, der sekundären und der tertiären (s. Abb. 4).

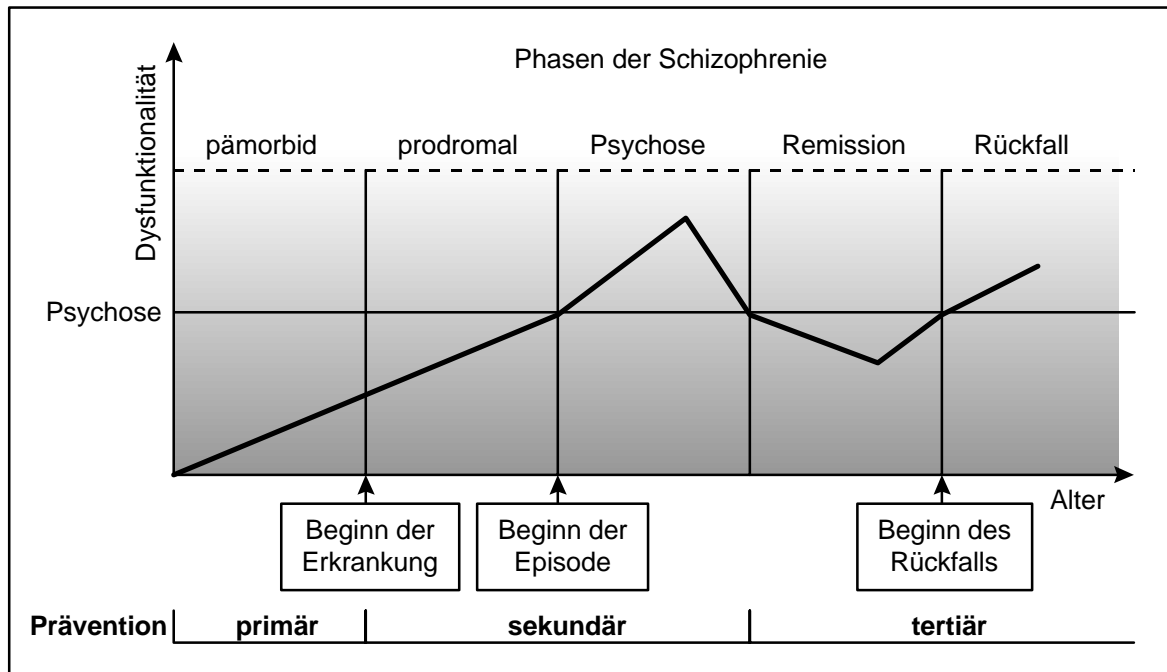


Abb. 4: Phasen einer schizophrenen Erkrankung und Präventionsstrategien (modifiziert nach Johannessen & Larsen, unpublished).

Eine primäre Prävention bzw. Prophylaxe schizophrener Erkrankungen, deren Erfolg in der Senkung der Inzidenz von Schizophrenie zum Ausdruck käme, wird nur durch unspezifische Interventionsformen in der Allgemeinbevölkerung, wie etwa verbesserte Bedingungen bei der Geburt zur Vermeidung von Geburtskomplikationen, Steigerung der allgemeinen Hygienestandards zur Vermeidung von Virusinfektionen (McGlashan, 1996) oder Beseitigung deprimierender Lebensbedingungen (van Os, Vortrag an der Psychiatrischen Klinik der Universität zu Köln, 12.08.1998), zu erreichen sein.

Dagegen umfaßt eine sekundäre Prävention gezielte Interventionsstrategien, die den Ausbruch einer akuten psychotischen Episode bei identifizierten Risikopersonen verhindern helfen sollen. Sie ist somit auf eine Reduktion der Prävalenz von Schizophrenie bzw. der Inzidenz von schizophrenen Erkrankungen im Sinne heutiger Diagnosekriterien nach ICD-10 und DSM-IV ausgerichtet. Demgegenüber zielt die tertiäre oder Rückfallprävention auf eine

Senkung der Morbidität ab und soll die Langzeitprognose bereits erkrankter Personen verbessern. Vor allem auf diesem Gebiet wurden in den letzten drei Jahrzehnten bereits erhebliche Erfolge erzielt.

So ergab hinsichtlich der Rückfallprävention ein Überblick über frühe Vergleichsstudien zum Verlauf bei neuroleptisch behandelten und neuroleptisch unbehandelten Patienten eindeutig, daß durch eine neuroleptische Medikation eine deutlich bessere Prognose zu erreichen war (Wyatt, 1991). Darüber hinaus konnte in neueren Studien gezeigt werden, daß eine weitere Verbesserung der Langzeitprognose über eine strukturierte Langzeitbehandlung mit zusätzlichen psychosozialen Interventionsansätzen erzielt werden konnte, die Familienintervention, supportive Psychotherapie und die Beobachtung früher Symptome eines Rückfalls, sogenannter *early signs* (Birchwood et al., 1989; Herz & Melville 1980), beinhalteten (Heinrichs, Cohen & Carpenter, 1985; Hogarty, Anderson & Reiss, 1986).

Die Identifikation von Prodromalsymptomen eines Rückfalls ist hierbei von besonderer Bedeutung, da ihr Auftreten oder ihre Zunahme den Zeitpunkt einer gezielteren Tertiärprävention vorgeben (Herz & Lamberti, 1995). Zur Erfassung der Prodromalsymptome eines Rezidivs wurden eine Reihe klinischer Modelle und Testverfahren entwickelt (Birchwood, 1992), wobei die bekanntesten sicherlich der 'Early Signs Questionnaire - ESQ' (Herz & Melville, 1980) und die 'Early Signs Scale - ESS' (Birchwood et al., 1989) sind. Eine Reihe von Studien zur Vorhersage schizophrener Rezidive demonstrierten dabei eine intraindividuelle Stabilität des interindividuell stark variierenden Musters von Prodromalsymptomen eines drohenden Rückfalls (Birchwood, 1992; Birchwood et al., 1989; Heinrichs & Carpenter, 1985; Herz & Melville, 1980), für die Birchwood (1992) den Begriff der *relapse signature* prägte.

Da oft angenommen wird, daß kein prinzipieller Unterschied zwischen tertiärer und sekundärer Prävention bzw. zwischen initialem und Rückfallprodrom besteht (Cutting, 1985; Docherty, van Kammen, Siris & Marder, 1978; Johannessen & Larsen, unpublished; Keith & Matthews, 1991), können die Ergebnisse zur Rückfallforschung bedingt auch auf das Prodrom einer schizophrenen Erstmanifestation ausgedehnt werden, wobei empirische Studien, die das Ausmaß der Übereinstimmung zwischen dem Prodrom eines Rezidivs und der Erstmanifestation in Einzelfällen untersuchen, bislang noch ausstehen. Ein Unterschied zwischen initialem und Rückfallprodrom kommt jedoch schon heute recht klar zum Ausdruck (Yung & McGorry, 1996): Prodrome eines psychotischen Rezidivs sind mit einer durchschnittlichen Dauer von einigen Wochen erheblich kürzer als das initiale Prodrom vor der psychotischen Erstmanifestation, das bis zu Jahrzehnten andauern kann. Mögliche Ursachen für eine solch unterschiedliche Dauer können zum einen in einer nach der ersten Episode erhöhten Vulnerabilität für eine psychotische Dekompensation, zum anderen aber auch

in einem reduzierten Schwellenwert für die Manifestation psychotischer Symptome nach dem Beginn einer psychotischen Störung liegen.

Im Gegensatz zu der Prävention von Rückfällen wird die Prävention einer ersten akuten schizophrenen Psychose in der Prodromalphase bei identifizierten Risikopersonen, die sekundäre Prävention, trotz der angenommenen Ähnlichkeiten zwischen Rezidiv- und initialen Prodromalsymptomen einer schizophrenen (Erst-)Episode erheblich dadurch erschwert, daß gerade die Identifizierung von Personen in einer schizophrenen initialen Prodromalphase auf Grund der im Vergleich zu der Akutsymptomatik größeren Unspezifität der Prodromalsymptomatik bis heute noch nicht mit hinlänglicher Sicherheit erfolgen kann (Malla, Norman & Voruganti, 1999; McGlashan, 1996; Vaglum, 1996; Yung & McGorry, 1997; Yung et al., 1996). Während diese Unspezifität bei der tertiären Prävention nur eine untergeordnete Rolle spielt - immerhin ist hier bereits bei dem einzelnen Patienten eine schizophrene Erkrankung bekannt, ist sie bei der sekundären Prävention von maßgeblicher Bedeutung, da sie die Gefahr in sich birgt, Personen, die noch niemals manifest schizophren erkrankt waren, fälschlich als 'prä-schizophren' oder in der Prodromalphase einer Schizophrenie befindlich zu diagnostizieren (Bell, 1992). Bevor also ähnliche Präventivmaßnahmen wie vor einem Rezidiv auch vor einer schizophrenen Erstmanifestation eingeleitet werden können, muß eine verlässliche Früherkennung schizophrener Psychosen gesichert sein.

2.1.5 Das Früherkennungsparadigma

Die Konzeption schizophrener Vulnerabilität, eingebettet in die Annahme eines schizophrenen Spektrums, geht von einer Normalverteilung psychotischer Züge in der Allgemeinbevölkerung aus, in der schizophren erkrankte Personen nur „the tip of an iceberg of human psychotic proclivity“ (McGlashan & Johannessen, 1996, p. 202) darstellen. So zeigte beispielsweise eine in Südfrankreich durchgeführte Feldstudie (van Os et al., 1999; Verdoux et al., 1998), daß paranoid wahnhaftige Überzeugungen nicht nur bei Personen mit schizophrener, schizoaffektiver oder affektiver Psychose vorkommen, sondern auch bei psychisch gesunden Patienten allgemeinmedizinischer Praxen auftreten.

Bei einer gezielten sekundären Prävention müssen also nicht nur Personen mit einer erhöhten Vulnerabilität für Schizophrenie identifiziert werden, sondern unter ihnen vor allem jene, die erste Anzeichen einer beginnenden schizophrenen Erkrankung bieten. So wird entsprechend im Vulnerabilitäts-Streß-Bewältigungs-Modell von Nuechterlein (Nuechterlein & Dawson, 1984) zwischen sogenannten *state*- und *trait*-Merkmale unterschieden. Während letztere zeitlich überdauernde Merkmale beschreiben, die eine generelle Anfälligkeit für schi-

zophrene Erkrankungen anzeigen, sind state-Merkmale zeitlich auf die schizophrene Episode begrenzt und kehren während der klinischen Remission auf einen Ausgangswert zurück, der sich nicht von dem in der Allgemeinbevölkerung unterscheidet. Sie sind somit kein Charakteristikum einer schizophrenen Vulnerabilität. Kennzeichnend für die Prodromalphase einer Schizophrenie werden Symptome betrachtet, die sowohl trait- als auch state-Charakter besitzen (Klosterkötter, 1996; Nuechterlein & Dawson, 1984; Parnas, 1999; s. Abb. 5).

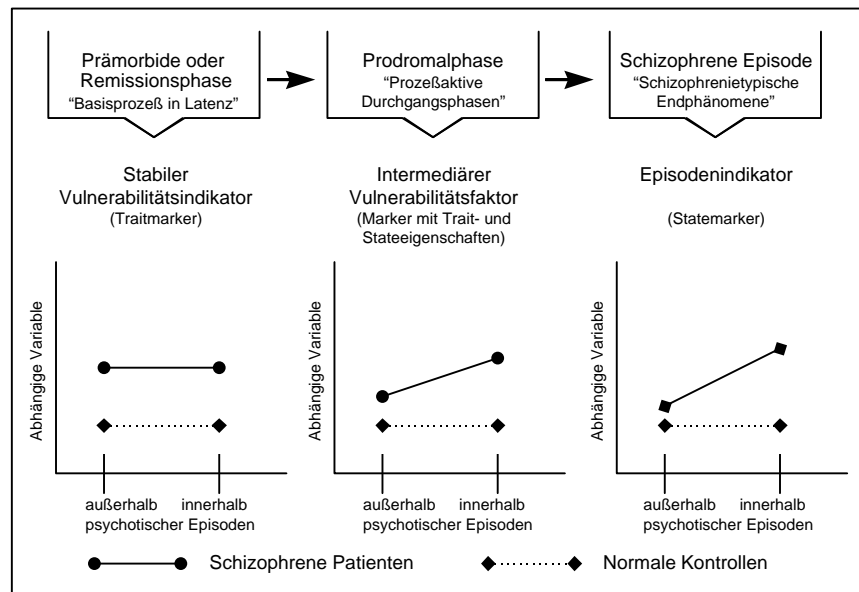


Abb. 5: Trait-state-Kontinuum (modifiziert nach Nuechterlein & Dawson, 1984)

2.1.5.1 Vulnerabilitätsindikatoren

Vulnerabilitätsindikatoren hingegen wird trait-Charakter zugeschrieben (s. Abb. 5). Sie werden deshalb überwiegend aufgrund von übereinstimmenden Auffälligkeiten bei bereits schizophren erkrankten Patienten und ihren (noch) psychisch gesunden biologischen Angehörigen oder aufgrund von Verlaufsbeobachtungen schizophrener Patienten im frühen Stadien der schizophrenen Erkrankung beschrieben (Cornblatt, Obuchowski, Schnur & O'Brien, 1998; Parnas, 1999). Die Identifikation solch überdauernder Vulnerabilitätsindikatoren wird als wesentlich bei der Aufdeckung eines schizophrenen Genotyps angesehen (Nuechterlein et al., 1992).

Tab. 1: Mögliche Vulnerabilitätsindikatoren der Schizophrenie zusammengestellt nach Übersichten von Carpenter und Buchanan (1994), Erlenmeyer-Kimling und Cornblatt (1987), Kremen und Mitarbeiter (1994), McGlashan und Johannessen (1996), Moldin und Erlenmeyer-Kimling (1994) sowie Nuechterlein und Mitarbeiter (1992).

<p>Klinisch Persönlichkeitsstörung des DSM-III-R-/DSM-IV-Clusters A Schizotypie im individuellen Fall oder in der Familie Familiäre Belastung mit einer Schizophrenie-Spektrum-Störung</p> <p>Verhaltensbezogen Frühe neurointegrative Defizite im Temperament, in der Erregung und in der Entwicklung Prämorbid auffälligkeiten: auf perzeptiv-kognitiven Defiziten basierende Verhaltensauffälligkeiten, affektiv, neuromotorische Dysfunktionen, geringe emotionale Ausdrucks- und Resonanzfähigkeit, soziale Kompetenzdefizite, Einzelgängertum, Antriebsschwäche, schulische und funktionelle Defizite</p> <p>Umweltbezogen Psychosozialer Stress: negatives affektives Klima, niedriger sozioökonomischer Status, instabiles Erziehungsumfeld Perinatale Faktoren: Wintergeburten, Influenza, Autoimmunschwäche, Hunger, RH Inkompatibilität, Schwangerschafts- und Geburtskomplikationen</p> <p>Anatomisch / neuroanatomisch Geringe physische Auffälligkeiten Fluktuierende anatomische Asymmetrien Strukturelle Hirnabweichungen: Erweiterungen der lateralen Ventrikel und der Basalganglienkerne, Veränderungen des limbischen Systems, des Gyrus parahippocampalis und des Gyrus cingulus anterior, Volumenverkleinerung der Amygdala, des Temporallappens und des Hippocampus Verringerung der Stoffwechselaktivitäten und des Blutfluß' im Frontallappen, Präfrontallappen und inferioren Parietallappen.</p> <p>Chemisch Homovanillinsäure im Blutplasma und der Zerebrospinalflüssigkeit bei einer schizotypischen Persönlichkeitsstörung Monoaminoxidase in Plättchen bei einer schizotypischen Persönlichkeitsstörung Dopaminerge Hypoaktivität im mesokortikalen verbunden mit Hyperaktivität im mesolimbischen System Veränderungen der norephedrinergen, serotonergen oder glutaminergen Aktivität</p> <p>Genetisch Polygenetischer Erbgang: bisher keine verlässlichen Ergebnisse zu Anzahl, Heterogenität und Art involvierter Gene oder Transmissionsreihen. Vorläufige Daten weisen auf unterschiedliche Erbformen bei produktiven und negativen Symptomen hin.</p> <p>Psychophysiologisch Autonome Reaktionen: Störungen des elektrodermalen Aktivierungsniveaus</p> <p>Motorisch Defizite in langsamen Augenfolgebewegungen Visuelles Scannen / Fixieren Halteinduzierte Muskelspannung Neuromotorische Defizite (Mirror drawing)</p> <p>Perzeptiv-kognitiv Defizite in der Vigilanz (Continuous Performance Test) Defizite in der Antwortbereitschaft bei einfachen Reaktionstests Defizite der selektiven Aufmerksamkeit (Span of Apprehension Test) Defizite der frühen Informationsverarbeitung (10-letter forced-choice Span of Apprehension Test, degraded-stimulus Continuous Performance Test) Defizite in kognitiven, visuell-räumlichen und sensorisch-motorische Inhibierungs- und Aktivierungsprozessen der Informationsverarbeitung (startle reflex, prepulse inhibition, backward masking, Embedded Figures Test, negative priming, ereignisevozierte Potentiale, mismatch negativity, P300 Latenz) Defizite in der Diskrimination mehrdeutiger Stimuli Perzeptiv-motorische Geschwindigkeit (Trail Making Test, Stroop Test) Lockerung der Assoziationen (semantic priming) Hemisphärische Integration / Asymmetrie (dichotic listening, Händigkeit, Textverständnis und -wiedergabe)</p> <p>Neuropsychologisch Geringer IQ Störung von Abstraktionsvermögen / Konzeptbildung (Object-sorting Test, Wisconsin Card Sorting Test) Defizite in der mentalen Kontrolle / Enkodierung (Auditory and visual digit span test, dichotic listening) Störungen verschiedener Gedächtnisfunktionen Feine Störungen der expressiven und rezeptiven Sprache Dyslexie</p>
--

Tabelle 1 gibt einen Überblick über potentielle Vulnerabilitätsindikatoren. Die Frage der Spezifität der darin aufgeführten Indikatoren für Schizophrenie ist meist noch ungeklärt, ebenso wie ihre Charakterisierung als trait- oder state-trait-Merkmale.

Während es also eine Fülle von möglichen Vulnerabilitätsindikatoren gibt, ist der Weg, auf dem der schizophrene Genotyp in einen schizophrenen Phänotyp übergeht, noch ungeklärt. Einer der hypothetischen Übergänge in eine schizophrene Erkrankung führt über eine biologisch-genetisch determinierte Vulnerabilität zu einer Instabilität in der neuronalen Entwicklung während der Schwangerschaft (Lewis & Murray, 1987). Diese Instabilität kann in einer abnormen Neuroanatomie resultieren (Cannon et al., 1993) oder in einer erhöhten Anfälligkeit des Fötus für neuronale Schädigungen durch Stressoren wie etwa Influenza (Barr, Mednick & Munk-Jørgenson, 1990), Hungersnöte und damit einher gehende Unterernährung (Susser & Shang, 1992) oder Morbus haemolyticus fetalis (Hollister, Laing & Mednick, 1996) insbesondere innerhalb des zweiten Schwangerschaftstrimesters.

Bislang ungeklärt bleibt jedoch die Frage, warum eine genetisch determinierte Vulnerabilität für Schizophrenie meist nicht ihren phänotypischen Ausdruck vor der Adoleszenz findet. Eine mögliche Erklärung hierfür liegt in der Ausweitung der Hypothese einer genetisch erhöhten Anfälligkeit für Störungen der embryonalen neuronalen Entwicklung auf die Adoleszenz. Hierbei wird angenommen, daß die in der Adoleszenz stattfindende Eliminierung kortikal-kortikaler Synapsen, das sogenannte *synaptic pruning* (Hüttenlocher, 1979), über ihre normalen entwicklungsmäßigen Grenzen hinaus geht (Hoffman & McGlashan, 1993; Hoffman et al., 1995).

So unzweifelhaft eine biologisch determinierte Vulnerabilität bei allen offenen Fragen heute auch ist (McGlashan & Johannessen, 1996; Parnas, 1999), so ist sie nach Erkenntnissen einer finnischen Adoptionsstudie jedoch keine hinreichende Bedingung für die Entwicklung einer akuten schizophrenen Episode (Tienari et al., 1985). Vielmehr muß noch - wie bereits in Abbildung 1 und 2 dargestellt - mindestens ein psychosozialer Stressor hinzutreten, um zu einer psychotischen Dekompensation zu führen.

Die Anwesenheit eines oder mehrerer Vulnerabilitätsindikatoren rechtfertigt also noch keine sekundärpräventiven Maßnahmen, auch wenn das Vorliegen mehrerer Indikatoren theoretisch das Risiko einer falsch-positiven 'prä-schizophrenen' Diagnose verringert (Bell, 1992).

2.1.5.2 Prodromalsymptome

Nach dem Modell von Nuechterlein (s. Abb. 1 und 5) führt ein ständiger Anstieg der subklinischen Störungen - repräsentiert durch die intermediären 'states' - zu der Entwicklung von

Prodromalsymptomen. Eine Vielzahl von Prodromalsymptomen sind sogenannte *vermittelnde* (engl.: „mediating“) Vulnerabilitätsfaktoren mit trait- und state-Charakter, die mit hoher Wahrscheinlichkeit genetisch beeinflusst sind und von denen zu erwarten ist, daß sie mit einer abnorm hohen Häufigkeit in ‘high-risk’-Populationen auftreten. Sie werden als hilfreich bei dem Versuch angesehen, Personen mit einer beginnenden Schizophrenie zu identifizieren. Im Gegensatz zu den überdauernden Vulnerabilitätsindikatoren in Nuechterleins Modell gelten sie als Verbindungsglieder, die der psychotischen Symptombildung kausal näher stehen als die Vulnerabilitätsindikatoren (Nuechterlein et al., 1992). Dabei wird angenommen, daß Prodromalsymptome oft zu einer Erhöhung des Streßniveaus führen, vermehrt die Entwicklung protektiver Faktoren hervorrufen und temporär das Vulnerabilitätsniveau verändern (s. Abb. 1). Hypothetisch evoziert die Interaktion von erhöhter psychobiologischer Vulnerabilität und erhöhtem Streßniveau so einen kontinuierlichen Anstieg der Dysfunktionalität, also der Schwere der intermediären states und der Prodromalsymptomatik, es sei denn, die vorhandenen protektiven Faktoren stellen einen ausreichenden Schutz dar. Tun sie dies nicht, wird die Schwelle zur psychotischen Symptombildung letztendlich überschritten (s. Abb. 2), und das Individuum erlebt eine akute psychotische Episode, die oft von Einbußen in der sozialen und arbeitsbezogenen Leistungsfähigkeit begleitet ist (s. Abb. 1; Nuechterlein et al., 1992).

Der Begriff *Prodrom* (griech. προδρομος: Vorläufer) bezeichnet hierbei in der Medizin die frühen Symptome, die der charakteristischen Manifestation einer akuten und voll entwickelten Krankheit vorausgehen. Damit stellt das Prodrom im engeren Sinne ein retrospektives Konzept dar, das zu diagnostizieren erst nach der Ausbildung der endgültigen Krankheitssymptomatik möglich ist (Yung & McGorry, 1996). In der Psychiatrie bzw. Schizophrenieforschung definiert sich das Prodrom entsprechend über die Zeitspanne präpsychotischer Auffälligkeiten, die eine Abweichung von den vorhergehenden Erlebnis- und Verhaltensweisen einer Person darstellen. Hierbei ist das initiale Prodrom einer psychotischen Erstmanifestation von Prodromen psychotischer Rezidive zu unterscheiden, in denen es meist nicht nur zu einer erneuten Ausbildung im Einzelfall bereits bekannter prodromaler Auffälligkeiten im Sinne einer ‘relapse signature’ kommt, sondern auch zu einem Anstieg möglicherweise bestehender Residualsymptome. Während diese generelle Definition unter Klinikern und Forschern weitgehenden Konsens erzielt, gibt es doch Differenzen über die genaue Bestimmung der Zeitspanne, über die ein initiales Prodrom zu definieren ist. Während eine Möglichkeit darin besteht, das Selbsterleben des Patienten für die Beschreibung der Prodromdauer heranzuziehen, d.h. das Prodrom durch die Zeit zwischen den ersten subjektiv erlebten subklinischen Veränderungen und den ersten subjektiv erlebten psychotischen Symptomen zu beschreiben, besteht eine zweite Möglichkeit darin, die Beobachtungen von Bezugs-

personen zugrunde zu legen, d.h. das Prodrom über die Zeit zwischen den ersten von den Bezugspersonen beobachteten subklinischen Veränderungen und den ersten beobachteten psychotischen Symptomen zu definieren (s. Abb. 6). Ein alternativer Ansatz ist die Konsensbildung zwischen den Angaben des Patienten und denen seiner Bezugspersonen (Beiser, Erickson, Fleming & Iacono, 1993; Häfner et al., 1990; Loebel et al., 1992), d.h. die Definition des Prodroms als Zeitspanne zwischen den validen und in den Angaben von Patient und Bezugspersonen möglicherweise übereinstimmenden Schilderungen erster subklinischer Veränderungen und erster psychotischer Symptome. Dabei können die Zeitangaben zu dem Beginn des Prodroms und dem der psychotischen Episode aufgrund der oft schleichend-graduellen Veränderungsprozesse meist nicht auf einen fixen Zeitpunkt - wie es in Abbildung 6 den Anschein hat, sondern nur auf ein Zeitintervall eingegrenzt werden. Diese Einschränkung gilt insbesondere für den Beginn des initialen Prodroms (Beiser et al., 1993). Eine Ursache hierfür liegt in der der Prodromdefinition innewohnenden Retrospektivität der Angaben, die abhängig ist von der Genauigkeit der Erinnerungen der Befragten.

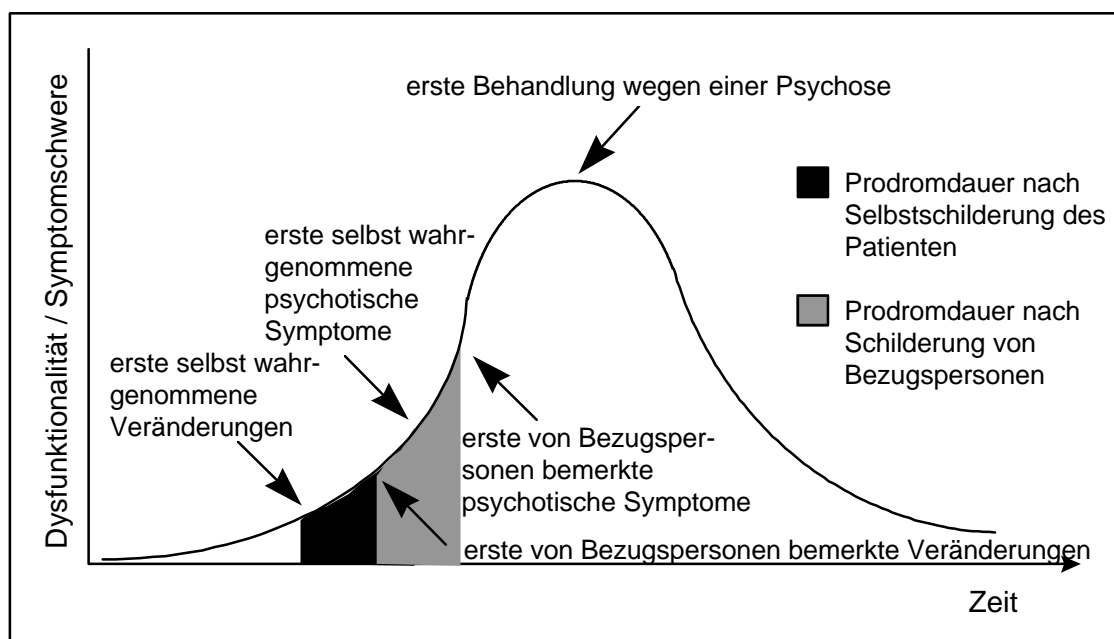


Abb. 6: Zeitpunkte der selbst- oder fremdbeobachtete Veränderungen bei der Entwicklung einer Psychose (modifiziert nach Yung & McGorry, 1996)

Es konnte gleichwohl gezeigt werden, daß auch rein subjektiv erlebte und bereits länger zurückliegende subklinische Veränderungen aus der Erinnerung des Betroffenen heraus noch zuverlässig zu ermitteln sind (Cutting & Dunne, 1989), und dies auch unabhängig vom schulischen Ausbildungsniveau (Jaeger, Bitter, Czobor & Volavka, 1990) oder dem allgemeinen Intelligenzniveau (Schultze-Lutter & Klosterkötter, 1995). Empirische Untersuchungen zu

dem möglichen Einfluß von individuellen Erklärungsmodellen für die spätere Psychose, von etwaigen Schuldgefühlen auf Seiten der Angehörigen, von Verdrängungsmechanismen als Bewältigungsstrategie der Psychose oder von dem psychisch-mentalenen Zustand des Patienten zum Zeitpunkt der Befragung stehen bislang noch aus. Um jedoch den Einfluß psychotisch bedingter kognitiver Denk- und speziell Gedächtnisdefizite im Sinne einer konfundierenden Störvariable zu vermeiden, sollte die retrospektive Erfassung präpsychotischer subjektiver Veränderungen in der Remissionsphase nach Abklingen der akut-psychotischen Symptomatik erfolgen (Maurer & Häfner, 1995).

Trotz der retrospektiven Definition eines Prodroms und der daraus resultierenden methodischen Probleme bei der Erfassung von Prodromalsymptomen wurde immer wieder auf die Bedeutung einer präventiv ausgerichteten Früherkennung und -behandlung von schizophrenen Psychosen zur Vermeidung schwerwiegender psychologischer und sozialer Folgeschäden hingewiesen (Birchwood & Macmillan, 1993; Cameron, 1938; Docherty, van Kammen, Siris & Marder, 1978; Falloon, 1992; Klosterkötter, 1998; Malla & Norman, 1994; McGorry, 1998; Parnas, 1999; Yung & McGorry, 1997). Doch um dies zuverlässig leisten zu können,

...a detailed characterization of the symptoms and signs of initial psychotic prodrome and a study of the evolution of prodromal symptoms to psychotic symptoms are necessary for early diagnosis and intervention. In particular, specific symptoms heralding a psychotic episode need to be identified. (Yung & McGorry, 1996, p. 356)

Mit anderen Worten wird also die Erstellung valider und reliabler symptom- und entwicklungsbezogener Vorhersagemodelle für psychotische Erstmanifestationen als Voraussetzung für eine Früherkennung und -behandlung angesehen.

2.1.5.3 Methoden zur Erfassung von Prodromalsymptomen

Bei der Erfassung von Prodromalsymptomen wurden neben den wenigen Geburtskohortenstudien (eine Übersicht geben Jones & Tarrant, 1998) bislang hauptsächlich vier verschiedene methodische Ansätze beschrieben (Yung & McGorry, 1996):

1. die detaillierte retrospektive Beschreibung von Veränderungen der prämorbidem Persönlichkeit des Patienten über erste prodromale Symptome bis hin zur Erstmanifestation einer Psychose, die durch Befragungen von Patienten in der Remissionsphase, ihrer Bezugspersonen und/oder anderer Informationsquellen wie etwa Studium der Krankenakten vorbehandelnder Ärzte gewonnen werden,
2. die Befragung von Patienten im frühen Stadium einer akuten Psychose und ihrer Bezugspersonen über aktuell erlebte Symptome,

3. prospektive Langzeituntersuchungen von Personen mit einem möglichen schizophrenen Prodrom oder einem aufgrund familiärer Belastung erhöhten Erkrankungsrisiko ('high-risk'-Studien),
4. die prospektive oder retrospektive Studie von Prodromen eines Rezidivs bei bereits diagnostizierten schizophrenen Patienten als Modell für das Prodrom einer schizophrenen Erstmanifestation.

Tabelle 2 gibt eine Übersicht über methodische Vor- und Nachteile dieser vier Vorgehensweisen.

Tab. 2: Methodische Vor- und Nachteile verschiedener Vorgehensweisen zur Erfassung von Prodromalsymptomen einer schizophrenen Erstmanifestation.

	1.*	2.*	3.*	4.*
abhängig von Zeit-, Gedächtnis- und Bewältigungsfaktoren	+	-	-	±
diagnostische Unsicherheit bei Erstdiagnosen (Amin et al., 1998; McGorry, 1994)	+	+	+	-
mögliche Beeinflussung durch aktuelle psychotische Positivsymptomatik	-	+	-	-
mögliche Beeinflussung durch aktuelle psychotische Negativsymptomatik	+	+	-	+
Stichprobenauswahl abhängig von Vorannahmen über Prodrome bei anfallenden klinischen Stichproben mit Verdacht auf ein psychotisches Prodrom	-	-	+	-
große Stichproben nötig bei ' high-risk '- Studien, um ausreichende Anzahl von später erkrankenden Personen, d.h. des Zielkriteriums, zu gewährleisten	-	-	+	-
Jahrzehnte umfassender Beobachtungszeitraum bei ' high-risk '-Studien	-	-	+	-
Gefahr möglicher Konfusion zwischen trait- und state-Merkmalen bei ' high-risk '-Studien	-	-	+	-
keine Aussage über Prodrome bei Personen ohne genetisches Risiko bei ' high-risk '-Studien	-	-	+	-
Gefahr hoher drop-out-Raten	-	-	+	±
keine Aussagen möglich über das Vorkommen von Prodromalsymptomen bei Personen mit beginnender Schizophrenie aber guten Bewältigungsstrategien und/oder gutem protektiven Umfeld, die die Schwelle zur psychotischen Symptombildung nicht überschreiten	+	+	-	+
Aussagen über Spezifität der erhobenen Symptomatik für Schizophrenie nur bedingt möglich	+	+	-	+
unbekannter Einfluß von Medikation und Behandlung	-	-	-	+
nicht empirisch abgesicherte Vorannahme über Zusammenhänge von Prodromen vor einer Erstmanifestation und Rezidiven	-	-	-	+

* die Numerierungen beziehen sich auf die Aufzählung der verschiedenen Studiendesigns (s.o.)

+ : zutreffend; - : nicht zutreffend; ± : bedingt zutreffend

Vom theoretischen Standpunkt aus gesehen sind sicherlich Populationsstudien wie die Geburts-Kohortenstudien, bei denen Probanden aus der Gesamtbevölkerung von Geburt an systematisch und prospektiv über den weiteren Verlauf ihres Lebens - insbesondere über den Zeitraum eines erhöhten Erkrankungsrisikos - beobachtet werden, die Methode der Wahl. Solche Studien bedürfen aber einer hohen Inzidenzrate der Erkrankung, da sonst der Stichprobenumfang enorm sein muß, um eine ausreichend große Anzahl von später erkrankenden Individuen einzuschließen. Die jährliche Inzidenzrate von Schizophrenien läßt ein solches Vorgehen mit schätzungsweise 10 bis 15 Erkrankungsfällen pro 100 000 Einwohnern deshalb nicht oder nur wenig praktikabel erscheinen (Maurer & Häfner, 1995).

Als ein praktikables Vorgehen zur Früherkennung wird heute die prospektive Studie von Personen mit einem vermutlich vorliegenden schizophrenen Prodrom gesehen (Vorgehen 3. in Tab. 2), wobei Prodrome anhand bisher vorliegender Ergebnisse aller vier Untersuchungsansätze theoretisch definiert werden, um sie einer empirischen Untersuchung insbesondere im Hinblick auf ihre positive prädiktive Stärke und Spezifität für schizophrene Erkrankungen zu unterziehen (Yung & McGorry, 1997). Tabelle 3 gibt eine Übersicht der in der Literatur als schizophrene Prodromalsymptomatik beschriebenen selbst wahrgenommenen Symptome und beobachtbaren Verhaltensweisen nach einer Zusammenstellung von Yung und McGorry (1996) wieder. Eine detaillierte und umfassende Beschreibung von selbst wahrgenommenen Beschwerden im Vorfeld schizophrener Erkrankungen und von Übergangsreihen in schizophrene Erstrangsymptome bietet vor allem das Basissymptom-Konzept (Huber, 1986; Klosterkötter, 1992), das in einem eigenen Kapitel behandelt wird (vgl. Kap. 2.1.7).

2.1.5.4 Entwicklungsmodelle psychotischer Erstmanifestationen

Neben einer möglichst genauen und umfassenden Darstellung von selbst- und fremdwahrgenommenen Symptomen im Prodromalstadium einer Schizophrenie ist für eine verlässliche Früherkennung auch eine möglichst zutreffende Beschreibung der Entwicklungsschritte von den ersten prodromalen Symptomen bis hin zu akut-psychotischen Symptomen wichtig, da Prodrome einen prozeßhaften Charakter haben und adäquat nicht durch eine bloße Auflistung von Symptomen zu einem beliebigen Zeit- bzw. Entwicklungspunkt der Prodromalphase beschreibbar sind. So beklagten schon 1978 Docherty und Mitarbeiter die 'statische' Ausrichtung psychiatrischer Forschung auf zu bestimmten Zeitpunkten vorhandene Symptome unter Vernachlässigung des prozeduralen Charakters psychischer Erkrankungen:

Tab. 3: Prodromalsymptome einer schizophrenen Erkrankung (nach einer Literaturübersicht von Yung & McGorry, 1996)

Symptombereich	Symptom
Neurotisch anmutende Symptome	<ul style="list-style-type: none"> • Ängstlichkeit * • Unruhe • erhöhte Reizbarkeit • erhöhte Irritierbarkeit *
Affektbezogene Symptome	<ul style="list-style-type: none"> • depressive Stimmungslage * • Anhedonie • Schuldgefühle • Mißtrauen * • suizidale Gedanken • Stimmungsschwankungen • allgemeinere Affektveränderungen
Veränderung der Willensstärke	<ul style="list-style-type: none"> • Apathie • Verlust des Antriebs und der Motivation * • Gefühle von Langeweile und Verlust von Interessen • Müdigkeit und Verlust von Energie *
Kognitive Denkstörungen	<ul style="list-style-type: none"> • Aufmerksamkeitsstörungen * • Konzentrationsstörungen * • vermehrte gedankliche Beschäftigung mit einem Thema, vermehrte Tagträume • reduziertes Abstraktionsvermögen (Konkretismus) • Denkblockaden
Körperliche Symptome	<ul style="list-style-type: none"> • somatische Beschwerden • Gewichtsverlust • Appetitlosigkeit • Schlafstörungen *
Andere Symptome	<ul style="list-style-type: none"> • zwanghafte Phänomene • dissoziative Phänomene • erhöhte interpersonelle Beeindruckbarkeit und Empfindsamkeit • Veränderungen in der Wahrnehmung der eigenen Person, anderer Personen und/oder der Welt im allgemeinen • motorische Störungen • Sprachstörungen • Wahrnehmungsstörungen
Beobachtbare Verhaltensänderungen	<ul style="list-style-type: none"> • Einbrüche in der Bewältigung von schulisch-beruflichen Anforderungen und in der sozialen Rollen Anpassung * • sozialer Rückzug * • impulsives Verhalten • ungewöhnliches, bizarres Verhalten • aggressives, störendes Verhalten

* am häufigsten beschriebene Symptome in retrospektiven first-episode-Studien

For many reasons psychiatric research and practice tend to focus on the major syndromes as static and relatively independent structures. However, clinical practice teaches that major disorders do not arise de novo. Almost always there seems to be a prodromal history - a period of progressive psychological and biological dysfunction.

This is reflected in the common colloquial use of the term “nervous breakdown”. Over the years we have been recurrently impressed by the presence in almost every carefully taken history of an inexorably progressing decline in patients who suffer a psychotic decompensation. We have also be impressed by the appearance of a characteristic pattern in the way this decline unfolds. (Docherty, van Kammen, Siris & Marder, 1978, p. 420)

Docherty und Mitarbeiter (1978) generierten aufgrund eigener Beobachtungen und Sichtung der verfügbaren Literatur ein fünfstufiges Übergangsreihenmodell (s. Abb. 7), wobei sie zu Beginn der prodromalen Symptomatik das Auftreten unspezifischer, oft noch im subjektiven Erleben verbleibender, neurotisch anmutender Symptome beschrieben, denen stärker vom prämorbidem Erleben abweichende attenuierte psychotische Symptome folgten und die psychotische Dekompensation einleiteten. Erste beobachtbare Verhaltensveränderungen und Defizite in der sozialen Rollenanpassung können ihnen zufolge, z.B. anhand von Leistungseinbußen, bereits während der ersten Phase mit unspezifischer, hauptsächlich subjektiver Symptomatik beobachtet werden (s. Abb. 7).

Die Modellvorstellung einer sich beginnend mit unspezifischen Symptomen über zunehmend Schizophrenie-charakteristischere, attenuierte Symptombildung entwickelnden Psychose wird heute wohl am häufigsten vertreten (Yung & McGorry, 1996) und liegt auch dem Basis-symptom-Konzept zugrunde (vgl. Kap. 2.1.7). Hier wird sie jedoch um die Konzeption sogenannter *Vorpostensyndrome* erweitert.

Vorpostensyndrome sind definiert als phasenhaft abgegrenzte, vollständig remittierende, dem initialen Prodrom vorangehende präpsychotische Stadien, die in ihrer Symptomatik Prodromen gleichen, aber im Gegensatz zu ihnen nicht kontinuierlich in eine erste psychotische Episode übergehen. Sie werden als „Ausdruck ein und desselben Krankheitsgeschehens“ (Huber, 1995, S. 133) angesehen, das auch dem initialen Prodrom, der akuten schizophrenen Psychose und den ihr möglicherweise folgenden reinen Defizienzsyndromen zugrunde liegt, und nicht als Ausdruck eines prämorbidem persönlichkeitsbedingten Defizits im Sinne eines vorauslaufenden Defekts nach Janzarik (1968). Das beschwerdefreie Intervall zwischen einem solchen Vorpostensyndrom und dem initialen Prodrom variiert beträchtlich und kann zwischen nur einigen Monaten bis hin zu Jahrzehnten betragen (Huber, 1995). Trifft dieses Vorpostensyndrom-Modell zu, so wird eine wichtige Aufgabe der Früherkennung darin bestehen, zwischen einem selbständig remittierenden Vorpostensyndrom und einem behandlungsbedürftigen initialen Prodrom zu unterscheiden.

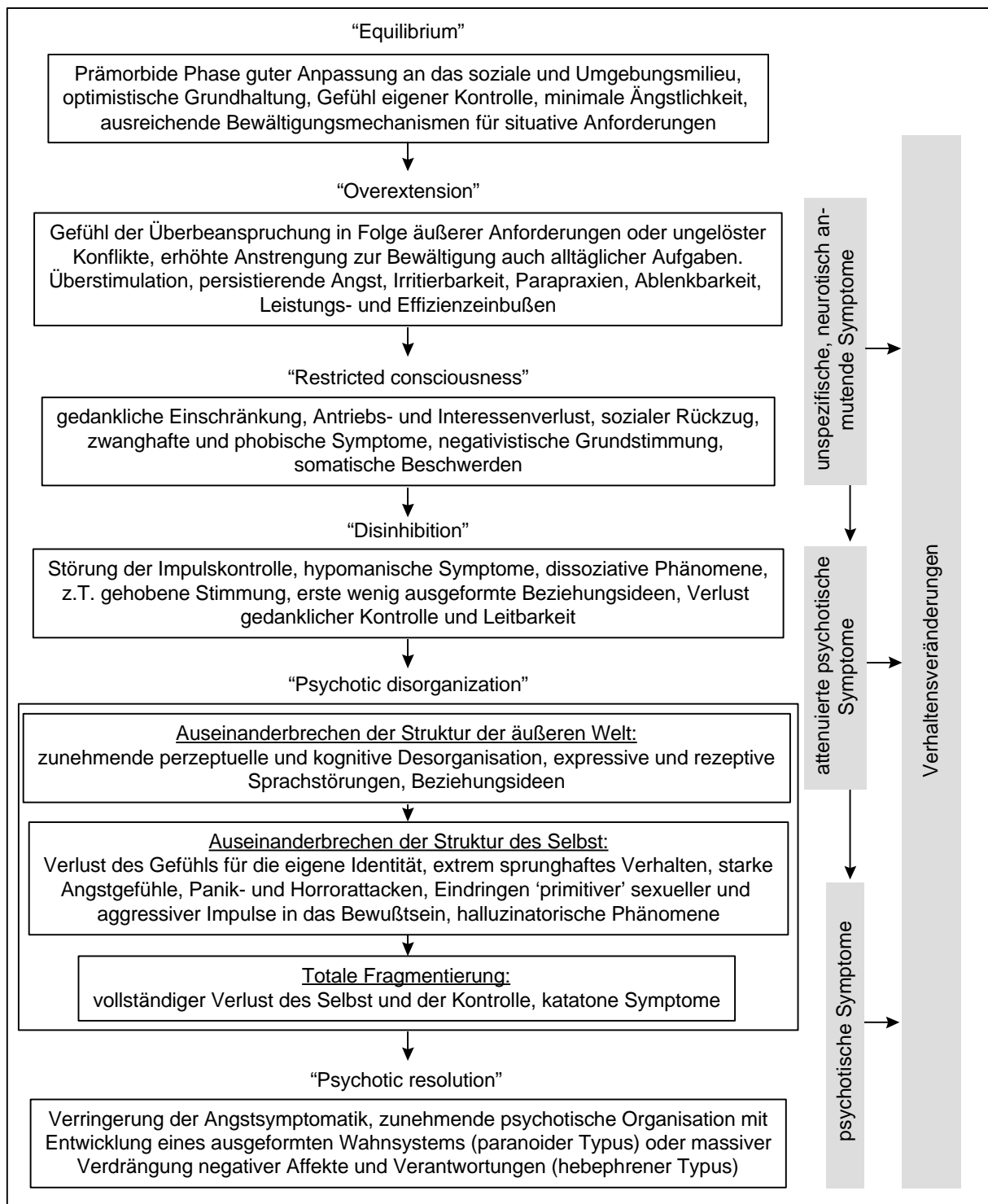


Abb. 7: Übergangsmodell nach Docherty et alii (1978). Unspezifischen Veränderungen folgen spezifische präpsychotische Symptome bevor es zur psychotischen Erstmanifestation kommt.

Eine alternative Modellvorstellung wurde von dem englischen Forscher James Chapman (1966, McGhie & Chapman, 1961) vertreten. Ihr zufolge entwickelt sich die floride psychotische Symptomatik direkt aus spezifischen Prodromalsymptomen, die vor allem durch selbst wahrgenommene kognitive Störungen der Aufmerksamkeit, der Wahrnehmung, der Handlungsausführung und von Denkprozessen gekennzeichnet sind (s. Abb. 8).

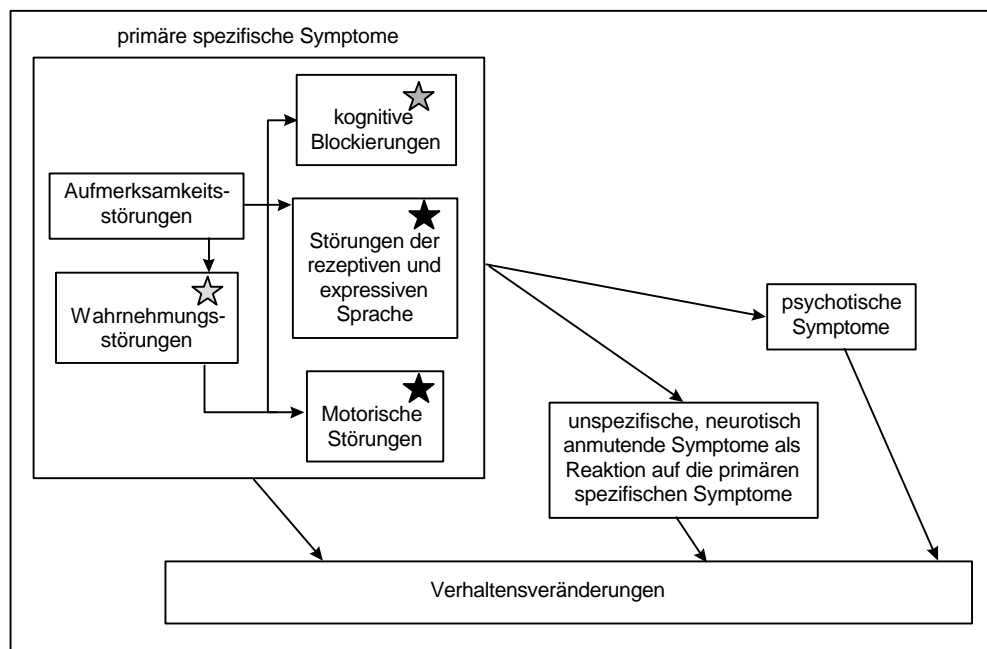


Abb. 8: Übergangsmodell nach Chapman (1966; McGhie & Chapman, 1961). Direkter Übergang von spezifischen Prodromalsymptomen in schizophrene Erstepisode. Unspezifische Symptomatik als Folge des Prodroms.

☆ Tönung indiziert Anteil anderer primärer Prodromalsymptome an der Entwicklung des entsprechenden Symptomkomplexes: hell=gering, dunkel=hoch.

Als primäre Störung wird hierin ein Defizit in den selektiven und inhibitorischen Steuerungsmechanismen der Aufmerksamkeitsfokussierung angenommen, in dessen Folge es zu einer graduell zunehmenden Regression („primitivization“, McGhie & Chapman, 1961, p. 114) von Wahrnehmungsprozessen, einer Diskriminationsschwäche zwischen sensorischen Kanälen, Störungen der motorischen Handlungskontrolle, einer erhöhten Wahrnehmung normalerweise unbewußter körperlicher Vorgänge sowie Konzentrations- und Denkstörungen kommt (s. Abb. 8). Unspezifische, neurotisch anmutende Störungen des Affekts und Antriebs treten nach dieser Modellvorstellung erst als Folge der erlebten primären Prodromalsymptome auf:

The main point we would wish to make here however is that when these affective changes are reported they would appear to be secondary and to follow as reactions to the patients awareness of the primary cognitive changes which have already taken place. (McGhie & Chapman, 1961, p. 110)

Zur Integration der drei unterschiedlichen Entwicklungsmodelle schizophrener Erstmanifestationen schlugen Yung und McGorry (1996) ein hybrides bzw. interaktives Modell vor (s. Abb. 9). Hiernach kann es im Vorfeld der ersten Episode einer schizophrenen Erkrankung zu einem Wechsel zwischen symptomatischen und asymptomatischen Phasen kommen, wobei

symptomatische Phasen von eher unspezifischen oder von eher spezifischen attenuierten Symptomen geprägt sein können. Sowohl unspezifische als auch attenuierte Symptome können der floriden Symptomatik direkt vorangehen oder primär bestehen, reaktive Beschwerdebilder wie Angst oder Depression sich sowohl aufgrund von prodromaler wie aufgrund von akut-psychotischer Symptomatik entwickeln und beobachtbare Verhaltensveränderungen durch jeden dieser Symptomkomplexe hervorgerufen werden.

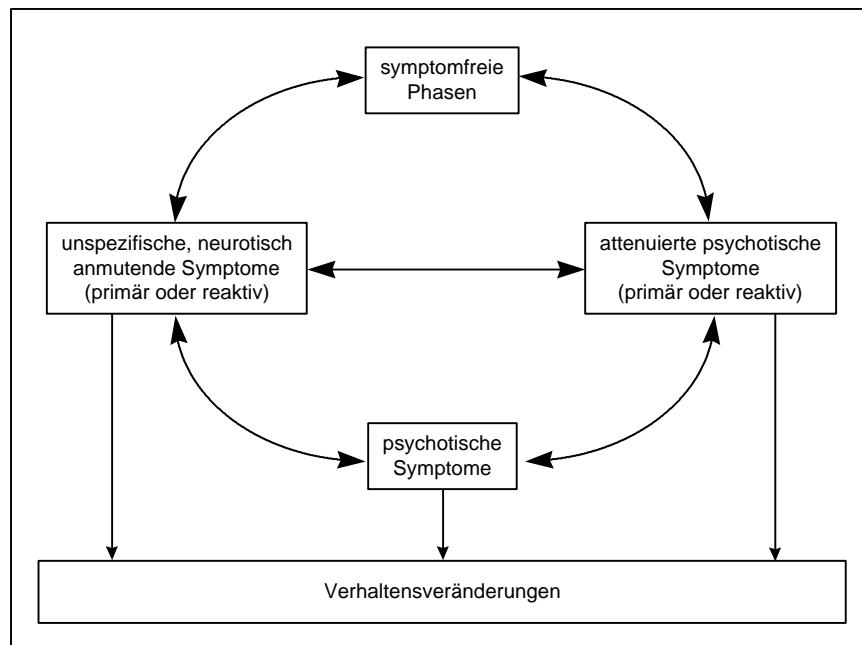


Abb. 9: Hybrides bzw. interaktives Übergangmodell nach Yung & McGorry (1996)

2.1.6 DSM-III-R-Prodromalsymptome einer schizophrenen Störung

Die beiden heute weltweit gebräuchlichsten Diagnosesysteme psychiatrischer Erkrankungen, ICD-10 (WHO, 1993) und DSM-IV (APA, 1994), bieten keine Richtlinien zur Diagnose eines schizophrenen Prodroms, obwohl in beiden die Prodromalphase als Bestandteil der schizophrenen Erkrankung erwähnt wird bzw. sogar als Teil der für die Diagnose notwendigen Erkrankungsdauer Berücksichtigung findet. Der Grund hierfür liegt in der geringen Spezifität in Frage kommender Symptome und der ungesicherten Reliabilität ihrer Erfassung (Jackson, McGorry & Dudgeon, 1995; Jackson, McGorry & McKenzie, 1994; Jackson et al., 1996; Keith & Matthews, 1991; McGorry et al., 1995).

Dennoch war zuvor im DSM-III und DSM-III-R (APA, 1980, 1987) der Versuch gemacht worden, prodromale Symptome einer schizophrenen Störung zu spezifizieren, die phänomeno-

logisch mit residualen Symptomen gleichgesetzt wurden und sich dadurch von diesen nur durch ihr Auftreten im klinischen Verlauf definitorisch unterschieden. Diese Kriterien sollten zu einer verbesserten Reliabilität beitragen (Andreasen & Flaum, 1991), obwohl sie nicht durch empirische Untersuchungen, sondern über einen Konsensentscheid des Komitees der American Psychiatric Association festgelegt wurden. Die Liste prodromaler bzw. residualer Symptome im DSM-III-R enthielt neun beobachtbare Verhaltensveränderungen, die eher atenuierte Formen positiver und negativer Symptome bzw. Merkmale der schizotypischen Persönlichkeitsstörung repräsentieren, jedoch keine subjektiv geschilderten Beschwerden:

1. soziale Isolierung oder Zurückgezogenheit
2. ausgeprägte Beeinträchtigung der Rollenerfüllung
3. ausgeprägt absonderliches Verhalten
4. ausgeprägte Beeinträchtigung der persönlichen Hygiene
5. abgestumpfter, verflachter oder inadäquater Affekt
6. abschweifende, vage oder umständliche Sprache
7. eigentümliche Vorstellungen oder magisches Denken
8. ungewöhnliche Wahrnehmungserlebnisse
9. erheblicher Mangel an Initiative, Interesse oder Energie

Für die DSM-III-R-Diagnose einer schizophrenen Psychose war das Vorliegen von mindestens zwei der neun Symptome vor der floriden Phase als Prodromalsymptome und/oder nach der floriden Phase als Residualsymptome über einen Zeitraum von insgesamt sechs Monaten notwendig, wobei in diesem Zeitraum das Auftreten einer akut-psychotischen Phase von mindestens einwöchiger Dauer mit florider Symptomatik obligat enthalten sein mußte, es sei denn, die psychotischen Symptome wurden erfolgreich behandelt.

Eine Arbeitsgruppe, die sich intensiv mit der Eignung der neun im DSM-III-R definierten Prodromalsymptome zur Früherkennung schizophrener Psychosen beschäftigt hat, ist die Melbourne Gruppe um Patrick D. McGorry und Henry J. Jackson (Jackson, McGorry & Dudgeon, 1995; Jackson, McGorry & McKenzie, 1994; Jackson et al., 1996; McGorry et al., 1995). Untersuchungen zur Spezifität und Prävalenz der DSM-Prodromalsymptome (Jackson, McGorry & Dudgeon, 1995; McGorry et al., 1995) an Patienten mit verschiedensten psychotischen Erstmanifestationen bzw. an durchschnittlich 16-jährigen high school-Schülern zeigten trotz höherer Prävalenzraten der Prodromalsymptome bei Patienten mit schizophrener Erstmanifestation, daß diese nicht nur kaum zwischen Patienten mit schizophrenen und anderen psychotischen Erstmanifestationen hinsichtlich ihrer Häufigkeit differierten, sondern auch bei psychisch unauffälligen high school-Schülern mit Prävalenzraten zwischen 8,1% für 'ausgeprägte Beeinträchtigung der persönlichen Hygiene' und 51,0% für 'eigentümliche Vor-

stellungen oder magisches Denken' vorlagen, wobei bei gut der Hälfte der 657 untersuchten Schüler mindestens zwei Symptome eruiert werden konnten.

Studien zur Interrater- und Test-Retest-Reliabilität (Jackson, McGorry & McKenzie, 1994; Jackson et al., 1996) an Patienten mit einer psychotischen Erstepisode ergaben insgesamt geringere Interrater- als Test-Retest-Reliabilitäten. Die Interrater-Reliabilität war zwar befriedigend ($0,31 \leq \kappa \leq 1,00$), wenn die Datenerhebung über ein Konsensverfahren zwischen Angaben des Patienten und dessen Bezugsperson unter Verwendung der detaillierten Itembeschreibungen und Entscheidungsregeln des 'Royal Park Multidiagnostic Instrument for Psychosis - RPMIP' (McGorry, Copolov & Singh, 1990; McGorry et al., 1990) erfolgte, jedoch teils inakzeptabel ($-0,06 \leq \kappa \leq 0,72$), wenn die Datenerhebung mit Hilfe des 'DSM-IV Field Trial Instruments - FTI' (APA, 1992) vorgenommen wurde, das keine operationalisierten Kriterien und Entscheidungsregeln enthält und sich ausschließlich auf den Patienten als Informationsquelle stützt. Die mit dem FTI erhobenen Test-Retest-Reliabilitäten erzielten demgegenüber mit einem κ zwischen 0,21 und 0,89 zufriedenstellendere Werte.

Die Melbournner Arbeitsgruppe (Jackson et al., 1996) zog aus ihren Untersuchungen den Schluß, daß die Streichung der DSM-III-R-Prodromalkriterien bei der Erstellung des DSM-IV durchaus berechtigt gewesen sei, da trotz bestehender Unklarheit über deren Validität die Erfassung zumindest bei Patienten mit schizophrener Erstmanifestation nur relativ unreliabel erfolgen könne. Gründe für die mangelnde Reliabilität lagen ihres Erachtens besonders in der unzureichenden Abgrenzbarkeit zu positiven und negativen Symptomen und in der Schwierigkeit, die Prodromalsymptome ausreichend von der Symptomatik bei schizoiden, paranoiden und insbesondere schizotypischen Persönlichkeitsstörungen zu unterscheiden. Aufgrund der unzweifelhaft bei den meisten schizophrenen Ersterkrankungen vorangehenden, lang andauernden Prodromalphase solle vielmehr statt der bisherigen beobachtbaren Prodromalsymptome nach einer alternativen Konzeptualisierung von Prodromen gesucht werden. Hinweise, wie solch alternative Prodromalsymptome definiert sein könnten, bestünden nach Ansicht der Autoren (ebd.) bereits in großer Zahl (s. Tab. 3): So haben etwa einige Forscher auf diesem Gebiet nicht-psychotische, neurotisch anmutende Symptome mit Augenmerk auf eine ängstliche und depressive Symptomatik vorgeschlagen (Birchwood et al., 1989; Malla & Norman, 1994), andere ziehen auch Symptome in Betracht, die als frühe Ausprägungen florider Symptomatik erscheinen (Carpenter & Heinrichs, 1983; Heinrichs & Carpenter, 1985; Herz, Glazer, Mirza, Mostert & Hafez, 1989).

Bevor nun aber Prodromalsymptome erneute Aufnahme in ein nosologisches System fänden, solle in breit gefächerten empirischen Untersuchungen an klinischen und allgemeinen Populationen geklärt werden, ob irgendeines der in Frage kommenden Symptome prädiktive

Validität besitze, ob es eventuell nur Ausdruck einer generellen Psychopathologie sei und/oder ob es möglicherweise spontan zurückgehe, ohne in eine psychotische oder andere psychische Erkrankung überzugehen. Hierfür seien v.a. prospektive Langzeituntersuchungen notwendig (Jackson et al., 1996). Die Autoren schließen ihre Ausführungen mit der Bemerkung:

Our view is that such other symptoms [wie sie alternativ zu der Definition im DSM-III-R von anderen Autoren vorgeschlagen wurden, v. Verf.], when taken in concert with one another, may have greater sensitivity and specificity, and more importantly, greater positive and negative predictive powers for psychosis, but obviously more empirical work needs to be undertaken as regards such symptoms. (ebd., p. 503f.)

Schritte in genau diese Richtung werden bereits seit einiger Zeit in Untersuchungen zum Basissymptom-Konzept unternommen.

2.1.7 Das Basissymptom-Konzept schizophrener Erkrankungen

Das Basissymptom-Konzept, ein integrativer Ansatz, der dem Vulnerabilitäts-Streß-Bewältigungs-Modell (Nuechterlein, 1987) vergleichbar ist, wurde in den 60er Jahren von Gerd Huber (Huber, 1966, 1983, 1986, 1995; Huber & Gross, 1989) entwickelt. Es hat seinen Ursprung in der Beobachtung von Defizienzen, die schon Jahre oder Jahrzehnte vor der ersten akuten Episode sowie im Vorfeld schizophrener Rezidive als auch postpsychotisch und intrapsychotisch bei fluktuierender akut-psychotischer Symptomatik auftreten, von den Betroffenen selbst wahrgenommen und (retrospektiv) berichtet werden (Huber, 1997; Huber, Gross & Schüttler, 1979; Huber, Gross, Schüttler & Linz, 1980). Diese milden, meist subklinischen, aber nichtsdestotrotz häufig starken Beschwerdedruck verursachenden Selbstwahrnehmungen von Störungen des Antriebs, des Affekts, der Denk- und Sprachprozesse, der Wahrnehmung, der Propriozeption, der Motorik und zentral-vegetativer Funktionen wurden von Huber unter dem Terminus technicus *Basissymptome* beschrieben, da er annahm,

...daß sie einem wie auch immer definierten somatischen Substrat, einem „hirnbedingten pathologischen Funktionswandel“ (*Conrad*), einer aktuellen Störung oder bleibenden Schädigung der seelischen Dynamik (50) [Janzarik, 1959; v. Verf.] näher sind als andere schizophrene Phänomene, etwa akzessorische Symptome von *E. Bleuler*, bei denen der „Prozeß“ nur die Disposition schafft, auf der psychodynamische Faktoren das Symptom entwickeln. (Huber, 1966, S. 415)

Basissymptome können als Teil der subjektiv wahrgenommenen Vulnerabilität aufgefaßt werden und sind mit objektiv meßbaren neuropsychologischen Defizienzen und, wenn auch

in geringerem Maße, mit der Intensität frontaler neurologischer Auffälligkeiten assoziiert (Cuesta, Peralta & Juan, 1996; Klosterkötter, Gross et al., 1997). Somit lassen sich das Basissymptom-Konzept und das Vulnerabilitäts-Streß-Bewältigungs-Modell leicht ineinander überführen (s. Abb. 10), indem die Basissymptome zum einen als Teil der psychometrisch erfaßbaren persönlichen Vulnerabilitätsfaktoren (s. Abb. 1) aufgefaßt werden, zum anderen den Platz der von Nuechterlein nur grob definierten Prodromalsymptome (s. Abb. 1) einnehmen (Klosterkötter, 1996). Diese Integration gelingt um so leichter, als zwar das Vulnerabilitäts-Streß-Bewältigungs-Modell auf Meehls Schizotypie-Konzeption (Meehl, 1990) basiert, das Basissymptom-Konzept hingegen auf E. Bleulers Vorstellung einer latenten Schizophrenie bzw. von formes frustes (Bleuler, 1911), beide zugrunde liegenden Konzeptionen jedoch von der Sache her den gleichen Tatbestand beschreiben (vgl. Kap. 2.1.1).

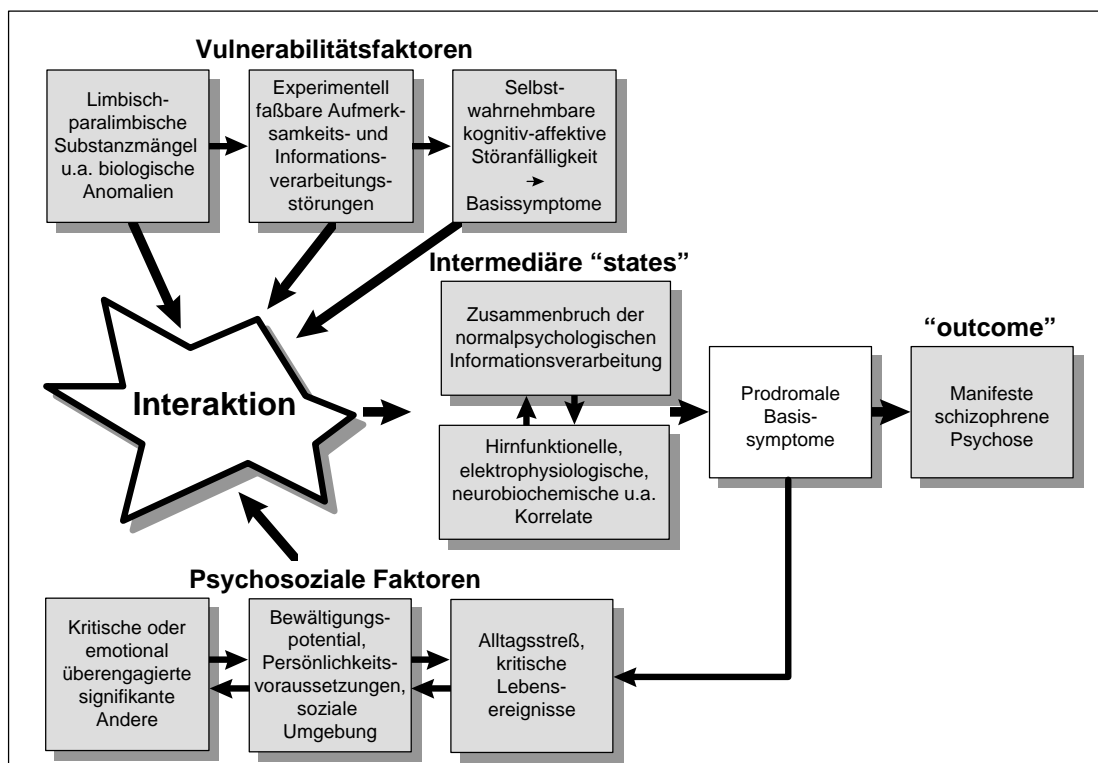


Abb. 10: Die Entwicklung schizophrener Psychosen nach dem integrativen Ansatz von Basissymptom-Konzept und Vulnerabilitäts-Streß-Bewältigungs-Modell (nach Klosterkötter, 1996)

Die Einbettung von Basissymptomen in die persönlichen Vulnerabilitätsfaktoren ist gerechtfertigt durch Nuechterleins Definition einer schizotypen Persönlichkeit im Sinne einer emotionalen Instabilität, wie sie in der 'high-risk'-Forschung beschrieben ist (Parnas & Mednick,

1991), eines subklinischen kognitiven Gleitens (Arboleda & Holzman, 1985) und feiner subjektiver Störungen von Denk-, Wahrnehmungs- und Propriozeptionsprozessen, wie sie mit den Wisconsin-Skalen (Chapman & Chapman, 1987), insbesondere der MAG, PER und PAN erhoben werden. Die hiermit beschriebene subjektive subklinische Symptomatik ist in ähnlicher Weise, obschon wesentlich detaillierter in der Operationalisierung des Basissymptom-Konzepts als Fremdbeurteilungsverfahren enthalten, der 'Bonner Skala für die Beurteilung von Basissymptomen - BSABS' (Gross, Huber, Klosterkötter & Linz, 1987, 1992a, 1992b, 1995; vgl. Kap. 3.1.2.2, Anhang A).

Die Ersetzung von Nuechterleins Prodromalsymptomen durch Basissymptome läßt sich ebenfalls mit der deutlich umfassenderen und präziseren Beschreibung möglicher prodromaler Symptome anhand der BSABS begründen, die nicht nur die Prodromalsymptome vor einem Rezidiv im Sinne der ESS und des ESQ sowie die subjektiven Korrelate der DSM-III-R-Prodromalsymptome erfaßt, sondern gerade in den Bereichen kognitiver Denk-, Wahrnehmungs- und Handlungsstörungen sowie ungewöhnlicher körperlicher Mißempfindungen, sogenannter *Coenästhesien*, detailliertere und darüber hinaus gehende Symptomdefinitionen enthält (Klosterkötter, 1996; Klosterkötter, Schultze-Lutter, Gross, Huber & Steinmeyer, 1997).

2.1.7.1 Entwicklungsmodell schizophrener Psychosen nach dem Basissymptom-Konzept

In der ursprünglichen Basissymptom-Konzeption wird von einem dreistufigen Prozeß bei der Entstehung einer schizophrenen Psychose ausgegangen: Über uncharakteristische, neurotisch anmutende *Stufe-1*-Basissymptome entwickeln sich schon leidlich charakteristische *Stufe-2*-Basissymptome, aus denen dann - sofern es sich nicht um ein spontan remittierendes Vorpostensyndrom, sondern um eine Prodromalphase handelt - die *Stufe-3*-Endphänomene der floriden psychotischen Symptomatik hervorgehen (Gross & Huber, 1989; s. Abb. 11). Die Basissymptome fluktuieren dabei hinsichtlich ihres Auftretens und ihrer Ausprägung nicht nur endogen, sondern auch in Abhängigkeit von äußeren Stressoren. Insbesondere eher charakteristische *Stufe-2*-Symptome sowie ihr Übergehen in eine floride Symptomatik können durch bestimmte Anlässe ausgelöst werden, etwa durch alltägliche soziale Situationen, alltägliche arbeitsmäßige oder ungewöhnliche, neue und unter Zeitdruck durchzuführende Beanspruchungen oder emotionale Minimalanlässe. Eine Provokation oder Verstärkung von Basissymptomen kann also prinzipiell durch jede beliebige Beanspruchung oder Situation erfolgen, die die individuelle krankheitsbedingt reduzierte Informationsverarbeitungskapazität überfordert.

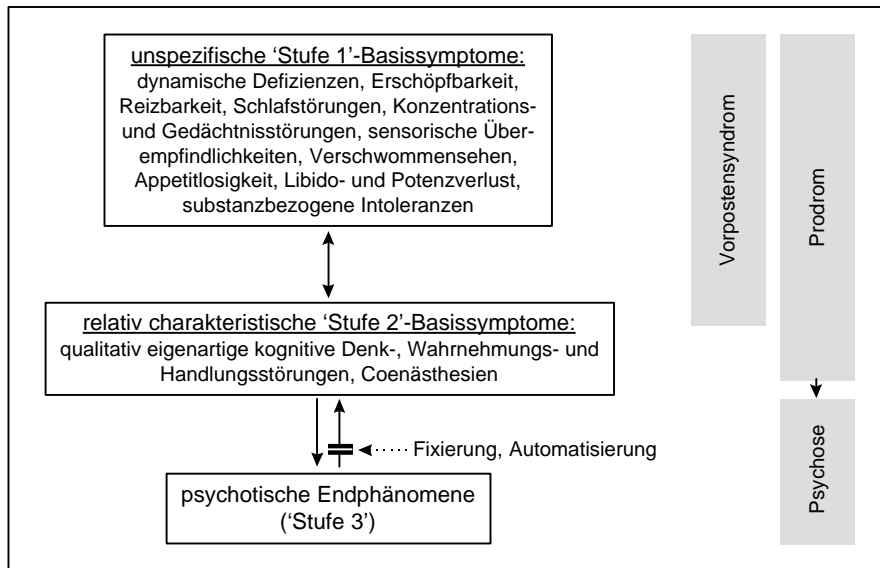


Abb. 11: Entwicklungsmodell schizophrener Psychosen nach der ursprünglichen Basissymptom-Konzeption.

Eine Umkehr von der dritten Stufe, der akuten Episode, über Stufe-2- und schließlich Stufe-1-Basissymptomen (sog. *postpsychotische reversible Basisstadien*) in die vollständige Remission ist möglich, so lange es nicht in der akuten Episode zu einer „möglicherweise im Zusammenhang mit einer prädisponierenden prämorbidem Persönlichkeitsstörung [stehenden; v. Verf.] Strukturverformung mit Fixierung und Automatisierung“ (Gross & Huber, 1989, S. 656) kommt (s. Abb.11). Bildet sich die Basissymptomatik jedoch im Anschluß an eine Psychose langfristig nicht zurück, so entsteht nach Gross und Huber (1989) ein *irreversibles reines Defizienzsyndrom* bzw. ein *reines Defektsyndrom*.

Eine Untersuchung von 121 Patienten mit der Diagnose einer paranoid-halluzinatorischen Schizophrenie nach ICD-9, die im Verlauf ihrer Erkrankung Erstrangsymptome, Defizienzerfahrungen im Sinne von Basissymptomen, zeitliche Folgen von solchen Basissymptomen und Erstrangsymptomen sowie darin eingebettete intermediäre Defizienzen geboten hatten, diente der gezielten Erhebung von Übergangsreihen zwischen vorauslaufenden subjektiven Defiziten und Erstrangsymptomen (Klosterkötter, 1992). Mit Ausnahme von 47 Patienten, die bei der Aufnahme eine anlaufende Basissymptomatik, aber noch keine akut-psychotischen Symptome zeigten und somit prospektiv untersucht werden konnten, wurden die Symptomabfolgen anhand einer freien Exploration und anschließender Untersuchung mit der neunten Version der 'Present State Examination - PSE9' (Wing, Cooper & Sartorius, 1974, 1982; vgl. Kap. 3.1.2.2) und der BSABS retrospektiv möglichst detailliert rekonstruiert. Auf diese Weise konnten 216 Symptomabfolgen dokumentiert und auf Übereinstimmungen hin analysiert werden. Hierbei war vor allem der Übergang von Stufe-2-Basissymptomen in die Erstrangsymptomatik von Interesse, weshalb sehr frühe Stufe-1-Symptome in der Analyse

vernachlässigt wurden. Es fanden sich charakteristische Übergänge für die Entwicklung von Wahrnehmungen, Gedankeneingebungs-, -entzugs- und -ausbreitungserlebnissen, akustische Halluzinationen ersten Ranges sowie für die Entwicklung von Willens- und leibliche Beeinflussungserlebnissen, denen zwar jeweils eine unterschiedliche Basissymptomatik zugrunde lag, die aber dennoch einem gleichen dreiphasigen Übergangsmuster folgten (s. Abb. 12).

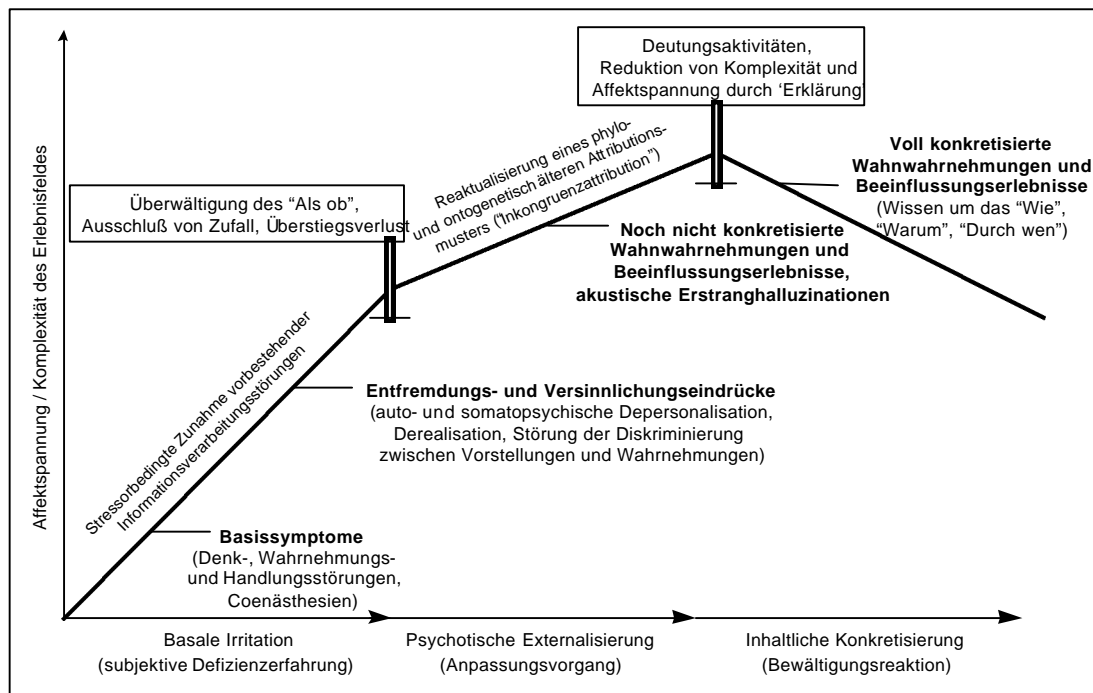


Abb. 12: Funktionale Betrachtung der drei Phasen der Entwicklung der Symptome ersten Ranges (modifiziert nach Klosterkötter, 1992)

Zu Beginn der psychotischen Entwicklung kam es zu einer zunehmend vom normalen Erleben abweichenden Basissymptomatik, der *basalen Irritation*, die meist durch äußere, die pathogen reduzierte Informationsverarbeitungskapazität überlastende Anforderungen hervorgerufen war. In dieser Phase trat noch keine psychotische Symptomatik auf, sondern der Betroffene erlebte seine gewohnte Erlebniswelt in wachsendem Maße als verändert und komplexer; seine Verwirrung und Affektspannung nahm zu. Der Bezug zur Realität, die sogenannte *Überstiegsfähigkeit* (Conrad, 1958), war in dieser Phase noch intakt; die erlebten Defizienzen besaßen 'Als ob'-Charakter.

Ein qualitativer und quantitativer Anstieg der Basissymptome führte im nächsten Schritt zu einem Zusammenbruch des normalpsychologischen Realitätssystems; es kam zu einer *psychotischen Externalisierung*. Für diese Phase wurden zwei Generierungsfaktoren postuliert: zum einen ein Überstiegsverlust durch Überwältigung des 'Als ob'-Charakters, zum anderen

eine Reaktivierung von phylo- und ontogenetisch älteren Realitätsbezugssystemen im Zuge eines Anpassungsversuchs an die als verändert erlebte Umwelt, das Zufälle ausschließt. Es kam zu der Überzeugung des 'Gemachten' der Erfahrungen, die im Sinne eines Wahneinfalls meist abrupt einsetzte. Diese leitete die Suche nach Erklärungsmodellen ein, die *inhaltliche Konkretisierung*. Hierbei entwickelte sich durch normalpsychologische Eigenaktivitäten, die durch die biographischen Erfahrungen und das zugängliche Wissen geprägt waren, schrittweise ein voll konkretisiertes Wahnsystems, in welchem das Wissen um die Gründe, Ursachen und Verursacher der erlebten Phänomene 'logisch' und Gegenargumenten unzugänglich eingebettet war. Dieses neue 'Wissen' führte zu einer Reduktion der Affektspannung, da durch die Erklärbarkeit die Komplexität der Erlebniswelt verringert wurde (s. Abb. 12). Tabelle 4 gibt die im ersten und zweiten Schritt der initialen Irritationsphase hauptsächlich gefundenen Basis- und PSE9-Symptome wieder, die am Beginn der Entwicklung der einzelnen Erstrangsymptome standen.

Doch so eindrucksvoll solche konsistenten Übergangsreihen auch sein mögen, reichen sie allein für die Frühdiagnose einer schizophrenen Erkrankung indes nicht aus, da hiermit noch nicht geklärt ist, wie häufig in der basalen Irritation auftretende Defizienzerfahrungen auch in nicht-schizophrenen Patientengruppen und in der Allgemeinbevölkerung vorkommen. Vielmehr bedarf es Untersuchungen zur Spezifität bzw. diagnostischen Validität solcher selbst wahrgenommenen präpsychotischen Defizienzen für eine schizophrene Erkrankung (Yung & McGorry, 1996).

2.1.7.2 Schizophreniespezifität von Basissymptomen

Vor allem die in den 80er Jahren durchgeführten Untersuchungen zur Schizophreniespezifität von Basissymptomen ließen Zweifel an der Eignung dieser Symptombilder für die Früherkennung aufkommen und führten dazu, daß Basissymptome oft als Ausdruck einer generellen und unspezifischen psychophysiologischen Desorganisation angesehen wurden, die sowohl bei Patienten mit kurzen psychotischen Episoden, neurotischen, hirnorganischen, substanz-induzierten, bipolaren und Zwangsstörungen als auch bei Strafgefangenen und Patienten mit Alkoholabhängigkeit in einer ähnlichen Häufigkeit auftreten wie bei Schizophrenen (Jimeno Bulnes, Jimeno Valdes & Conde Lopez, 1991; Kryspin-Exner & Lutterotti, 1982; Maß, Krausz & Gross, 1995; Mundt & Kasper, 1987; Ricca et al., 1997; Rösler, Bellaire, Henges, Kiesling-Muck & Carls, 1985; Teusch, 1984, 1985). Diesen Studien ist jedoch gemeinsam, daß sie alle anhand eines reinen Selbstbeurteilungsverfahrens zur Erfassung von Basissymptomen, des 'Frankfurter Beschwerdefragebogens - FBF' (Süllwold, 1977, 1986, 1991), durchgeführt wurden.

Tab. 4: Häufigste Defizienzerlebnisse am Anfang der Entwicklung von Erstrangssymptomen in der Reihenfolge der Häufigkeit (nach Klosterkötter, 1992). Die angegebenen Kennziffern beziehen sich auf die Itemnumerierung der BSABS (vgl. Anhang A).

Erstrangssymptom	Ausgangserfahrungen	Zwischenphänomene
Wahnwahrnehmung	Veränderungen an Gesicht / Gestalt anderer: C.2.3 s5 Mikro- und Makropsien: C.2.3 s2 Veränderung des Farbensehens: C.2.3 s4 akustische Qualitäts-, Intensitätsveränderungen: C.2.5 s1 Photopsien: C.2.2 s1 Scheinbewegungen von Objekten: C.2.3 s9 Spiegelphänomene: C.2.3 s6 gustatorische, olfaktorische Wahrnehmungsveränderungen: C.2.6 partielles Sehen: C.2.1 s3 Metamorphopsien: C.2.3 s3 sensorische Überwachtheit: C.2.8	Derealisation: C.2.11 Wahnstimmung Stufe 1: PSE9-Symptom Nr. 49 Wahnwahrnehmung Stufe 2: PSE9-Symptom Nr. 73
Gedankeneingebung, -ausbreitung, -entzug	Störung der Konzentrationsfähigkeit: C.1.5 Gedankeninterferenz: C.1.1 zwangähnliches Perseverieren: C.1.2 Gedankenblockierung: C.1.4 Gedächtnisstörungen: C.1.8 - C.1.11 Störung der Denkinitiative: C.1.13 Denkverlangsamung: C.1.12 Gedankendrängen / Gedankenjagen: C.1.3 Störung der expressiven Sprache: C.1.7 Störung der Diskriminierung von Gedanken und Wahrnehmungen: C.1.15	autopsychische Depersonalisation (v.a. gegenüber eigenen Denkabläufen): B.3.4 Wahnwahrnehmung Stufe 2: PSE9-Symptom Nr. 73
imperative, kommentierende und dialogische Phoneme	Störung der Konzentrationsfähigkeit: C.1.5 Gedankendrängen / Gedankenjagen: C.1.3 zwangähnliches Perseverieren: C.1.2 Gedankeninterferenz: C.1.1 Störung der Diskriminierung von Gedanken und Wahrnehmungen: C.1.15 Gedankenblockierung: C.1.4 Gedächtnisstörungen: C.1.8 - C.1.11 Störung der Denkinitiative: C.1.13 Störung der expressiven Sprache: C.1.7 Denkverlangsamung: C.1.12	Störung der Diskriminierung von auditiven Vorstellungen und akustischen Wahrnehmungen: C.1.15 Gedankenlautwerden: PSE9-Symptom Nr. 56
Willensbeeinflussung	Automatosesyndrom: C.3.1 motorische Interferenz: C.3.1 Bannungszustände: C.3.2 motorische Blockierung: C.3.2	autopsychische Depersonalisation: B.3.4 Wahnstimmung Stufe 1: PSE9-Symptom Nr. 49 Wahnwahrnehmung Stufe 2: PSE9-Symptom Nr. 73
leibliche Beeinflussungserlebnisse	Bewegungs-, Zug- und Druckempfindungen: D.7 mehr umschriebene Schmerzempfindungen: D.3 Elektrisierungssensationen: D.5 Sensationen der Verkleinerung, Schrumpfung und Einschnürung, der Vergrößerung und Ausdehnung: D.9 Taubheits- und Steifigkeitsempfindungen: D.1 Wandersensationen: D.4 Sensationen motorischer Schwäche: D.2 thermische Sensationen: D.6 Sensationen abnormer Schwere, Leichtigkeit und Leere, Fall-, Sink-, Levitations- und Elevationsphänomene: D.8	somatopsychische Depersonalisation: D.1 s1 Wahnstimmung Stufe 1: PSE9-Symptom Nr. 49 Wahnwahrnehmung Stufe 2: PSE9-Symptom Nr. 73

Obwohl bereits 1986 eine Synopse von FBF und BSABS vorgelegt worden war (Süllwold & Huber, 1986), gab es bis vor kurzem keine empirische Prüfung der Übereinstimmungen und Unterschiede der beiden Verfahren, die Aufschluß über eine Vergleichbarkeit der mit den beiden Verfahren gewonnenen Ergebnisse gegeben hätte. Eine 1997 vorgelegte Studie an 151 Patienten mit Diagnose einer Schizophrenie oder Alkoholabhängigkeit sowie der Doppeldiagnose beider Störungen läßt eine Vergleichbarkeit aber eher zweifelhaft erscheinen (Maß, Hitschfeld, Wall & Wagner, 1997). So fanden sich zwar hoch signifikante Korrelationen zwischen FBF und BSABS sowie deren Subskalen, doch gingen diese nicht über eine Korrelation von 0,60 zwischen der FBF-Subskala 4, expressive und rezeptive Sprachstörungen, und der BSABS-Kategorie C.1, kognitive Denkstörungen, die ebenfalls Störungen der expressiven und rezeptiven Sprache beinhaltet, hinaus ($0,28 \leq r \leq 0,60$). Berechnungen der Korrelation einzelner FBF-Items mit inhaltlich übereinstimmend erscheinenden BSABS-Items ergaben darüber hinaus selbst bei relativ eindeutigen, einfach definierbaren Symptomen im allgemeinen sogar noch schwächere Zusammenhänge, z.B. ergab sich für das Symptom 'Verschwommensehen' (FBF-Item 67, BSABS-Item C.2.1 s1) lediglich eine Korrelation beider Skalen von 0,51, d.h. Gemeinsamkeiten zwischen beiden Items bestanden nur zu etwa 25%. Die Autoren führen diese überraschend geringen Korrelationen auf mögliche unspezifische Kovarianzanteile zurück, wie etwa allgemeine Klagsamkeitstendenzen, Antwortstile (Schahn & Amelang, 1992) oder Neigung zur Krankheitsverleugnung, die in den beiden Untersuchungssettings, Beantwortung eines Selbstbeurteilungsbogens versus direktes Gespräch, in unterschiedlichem Maße zum Tragen kommen und kontrolliert werden können, obwohl beide Verfahren in der Hauptsache das Selbsterleben der Patienten erfassen.

So steht es nicht im Widerspruch zu Ergebnissen zum FBF, daß Studien mit der BSABS durchaus auf eine Schizophreniespezifität zumindest einiger Kategorien bzw. Subsyndrome hinweisen (Bechdorf, Halve, Schultze-Lutter & Klosterkötter, 1998; Ebel, Gross, Klosterkötter & Huber, 1989; Klosterkötter, 1994; Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996; Klosterkötter, Albers, Steinmeyer, Hensen & Saß, 1994; Klosterkötter, Schultze-Lutter et al., 1997, Klosterkötter, Gross et al., 1997; Schultze-Lutter & Klosterkötter, 1995). Die meisten dieser Vergleichsstudien sind neueren Datums und basieren auf einer BSABS-Kurzfassung von 66 Items (vgl. Anhang B), die seit Anfang der 90er Jahre an der Psychiatrischen Klinik der Rheinisch-Westfälischen Technischen Hochschule Aachen verwandt wurde. Diese Fassung stellt bereits eine zeitlich ökonomischere Variante der ursprünglich 98 Hauptitems (142 Einzelitems) umfassenden BSABS und nach damaligen Erkenntnissen getroffene Auswahl eher spezifischerer Basissymptome dar. Während die Kategorie der zentral-vegetativen Störungen (E) vollständig eliminiert wurde, wurden die Kategorien der dynamischen Defizienzen (A und B) teilweise gekürzt und zusammengefaßt, indem auf eine Unterscheidung von dy-

namischen Defizienzen mit direkter (A) und indirekter Minussymptomatik (B) verzichtet wurde. Die Kategorien der kognitiven Denk-, Wahrnehmungs- und Handlungsstörungen (C) sowie der Coenästhesien (D) fanden hingegen eine fast vollständige Aufnahme in die Kurzfassung.

Eine an 243 konsekutiven Aufnahmen der Psychiatrischen Klinik der RWTH Aachen (ICD-10-Diagnosen F0, F1, F2, F3, F4 und F6) sowie 79 psychisch gesunden Kontrollpersonen durchgeführte clusteranalytische Überprüfung der empirischen Itemgruppierung der Kurzfassung führte zu einer inhaltlich plausiblen und 53,25% der Varianz erklärenden Lösung von fünf Subsyndromen (Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996; s. Abb. 13):

- *Informationsverarbeitungsstörungen* (BIV): mit Ausnahme der 'Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten' (A.6.2) bestehend aus kognitiven Denk-, Wahrnehmungs- und Handlungsstörungen,
- *Coenästhesien* (BC): allen aus der Kategorie D, Coenästhesien, enthaltenen Items entsprechend,
- *Vulnerabilität* (BV): gebildet aus den fünf aus den Subkategorien A.8 und B.1, 'Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten Stressoren mit direkter oder indirekter Minussymptomatik', übernommenen Items,
- *Adynamie* (BA): eher kognitiv-dynamische Prozesse beschreibende Basissymptome aus der Subkategorie der Denkstörungen (C.1) - Konzentrationsstörungen, Störungen des Arbeits- und Kurzzeitgedächtnisses sowie Denkverlangsamung und Störung der Denkinitiative - als auch dynamische Affekt- und Kontaktstörungen charakterisierende Basissymptome (A.6) umfassend,
- *Interpersonelle Verunsicherung* (BIP): Anzeichen einer erhöhten Beeindruckbarkeit (B.2), Störungen der Kontaktfähigkeit und des In-Erscheinung-Tretens (A.7) sowie 'Eigenbeziehungstendenz' aus der Subkategorie der kognitiven Denkstörungen (C.1) einschließend.

Als besonders interessant bei dieser Gruppierung und gewissermaßen als eine Bereinigung der Kategorie der Denkstörungen (C.1) wurde deren Teilung in eher eigenartigere Stufe-2-Basissymptome und eher uncharakteristische, dynamische Denkstörungen angesehen, die in ein Cluster mit dynamisch-affektiven Störungen der Grundstimmung und der Kontaktbereitschaft eingehen. Auch die weitere Aufspaltung der verbleibenden dynamischen Defizienzen (A und B) in Symptome einer verringerten Belastungsfähigkeit gegenüber bestimmten von außen herangetragenen Anforderungen und Symptome einer verringerten emotionalen Belastungsfähigkeit bzw. erhöhten emotionalen Beeindruckbarkeit, die in Verbindung mit der Eigenbeziehungstendenz aus der Kategorie C.1 treten, wurde als inhaltlich plausibel betrachtet (Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996).

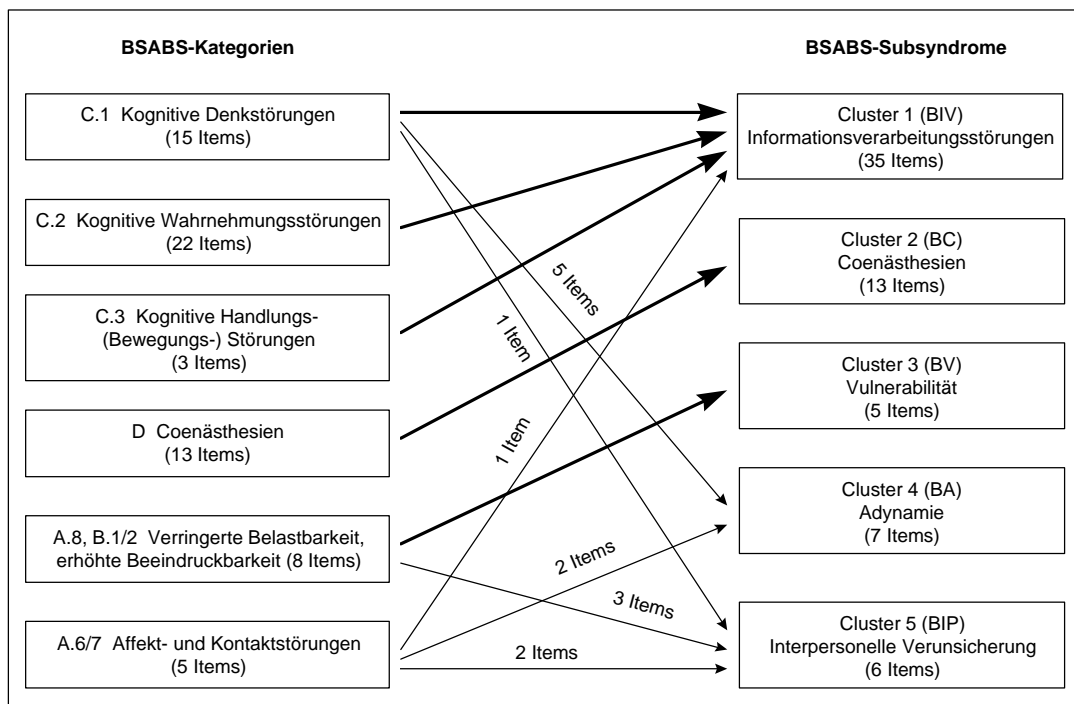


Abb. 13: Empirische Itemgruppierung der BSABS-Kurzfassung (nach Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996).

Eine anschließende Überprüfung dieser fünf Subsyndrome über die verschiedenen Patientengruppen und die Kontrollpersonen ergab für die interpersonelle Verunsicherung als einzigem Subsyndrom einen signifikanten Häufigkeitsvorsprung der schizophrenen Gruppe gegenüber allen anderen Gruppen (s. Abb. 14).

Im Gesamtwert und den Subsyndromen 'Informationsverarbeitungsstörungen' und 'Coenästhesien' konnte hingegen kein signifikanter Mittelwertsunterschied zwischen der schizophrenen und der hirnanorganischen Gruppe ermittelt werden, 'Adynamie' und 'Vulnerabilität' hingegen trennten nicht zwischen den drei Psychosegruppen, F0, F2 und F3 (s. Abb. 14). Kanonische Diskriminanzanalysen der Gruppen über die einzelnen Subsyndrome bestätigten die Ergebnisse der Mittelwertsvergleich weitgehend, ergaben jedoch für die 'Informationsverarbeitungsstörungen' im Gegensatz zu den 'Coenästhesien' eine hoch signifikante kanonische Komponente ($p < 0,0001$), die auch zwischen hirnanorganischen und schizophrenen Psychosen zu trennen in der Lage war (Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996).

Zu ähnliche Ergebnisse führte auch eine Vergleichsstudie zwischen chronisch Schizophrenen, leicht geistig Behinderten, Personen mit der Doppeldiagnose einer Schizophrenie und leichten geistigen Behinderung sowie geistig und psychisch unbeeinträchtigten Kontrollpersonen (Schultze-Lutter & Klosterkötter, 1995; s. Abb. 15). Hierin wurde die Hypothese geprüft, daß durch eine leichte geistige Behinderung vorbestehende, subjektiv erlebte und Basis-symptomen phänomenologisch gleichende und vor allem den affektiven und kognitiven

Bereich betreffende Defizite eine mögliche Erklärung für das erhöhte Schizophrenierisiko dieser Personengruppe von etwa drei Prozent (Turner, 1989) bieten könne, indem sie die Anfälligkeit bzw. Vulnerabilität für eine Schizophrenie bei Personen mit einer leichten geistige Behinderung unbekanntem Ursprungs erhöhen.

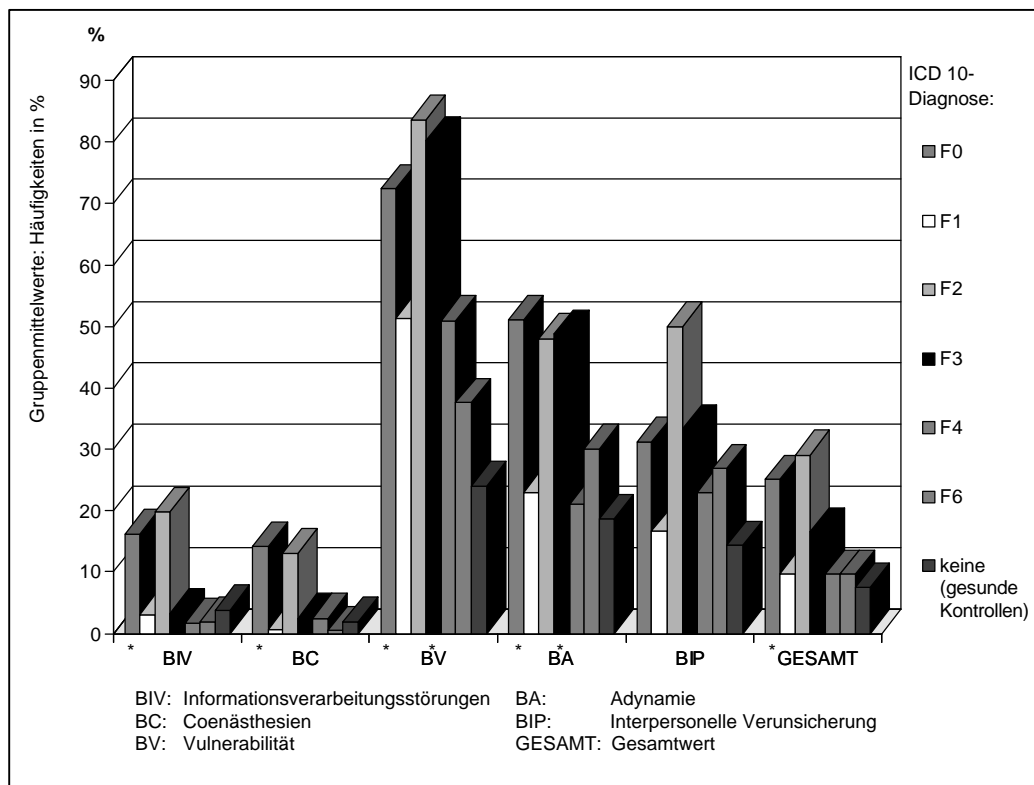


Abb. 14: Mittelwertsvergleich der BSABS-Subsyndrome über die sechs Patientengruppen und eine psychisch gesunde Kontrollgruppe, multiple t-Tests (nach Klosterkötter, Ebel, Schultze-Lutter & Steinmeyer, 1996). (F0: hirnorganische psychische Störungen; F1: Störungen durch psychotrope Substanzen; F2: schizophrene Störungen; F3: affektive Störungen; F4: neurotische, Belastungs- und somatoforme Störungen; F6: Persönlichkeitsstörungen) * $p > 0.01$ versus schizophrene Gruppe (F2): **kein signifikanter Unterschied**

Während sich Personen mit Schizophrenie unabhängig von ihrem intellektuellen Niveau erwartungsgemäß nicht voneinander, aber in jedem Fall von der Kontrollgruppe unterschieden, zeigten sich keine Mittelwertsunterschiede zwischen leicht geistig Behinderten und beiden schizophrenen Gruppen in den Subsyndromen 'Vulnerabilität' und 'Adynamie' (s. Abb. 15). Leicht geistig Behinderte hingegen berichteten im Vergleich zu der Kontrollgruppe ein hypotesenkonform allgemein erhöhtes Niveau von Basissymptomen phänomenologisch gleichenden Defiziten, wobei die Differenz im Gesamtwert sowie in den Subsyndromen 'Informationsverarbeitungsstörungen' und 'Interpersonelle Verunsicherung' signifikant wurde (Schultze-Lutter & Klosterkötter, 1995).

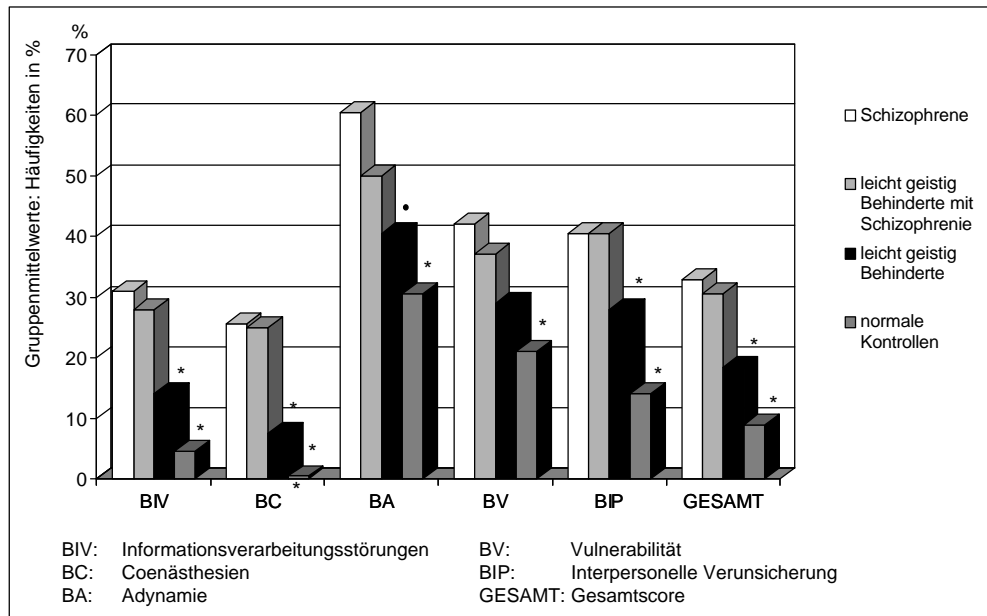


Abb. 15: Mittelwertsvergleich der BSABS-Subsyndrome über die vier Gruppen, multiple t-Tests (nach Schultze-Lutter & Klosterkötter, 1995).
 * $p < 0.01$ versus schizophrene und doppelt diagnostizierte Gruppe
 • $p < 0.01$ versus schizophrene Gruppe

Weitere Hinweise auf eine mögliche Schizophreniespezifität der Subsyndrome 'Informationsverarbeitungsstörungen' und 'Interpersonelle Verunsicherung' ergab eine Vergleichsstudie von Schizophrenen, psychisch gesunden erstgradigen Verwandten Schizophrener, die mit einem durchschnittlichen Alter von 27 Lebensjahren noch ein hohes Risiko für die Entwicklung einer schizophrenen Erkrankung besaßen, und familiär unbelasteten, psychisch gesunden Kontrollpersonen (Klosterkötter, Gross et al., 1997; s. Abb. 16). Die drei Gruppen waren im Hinblick auf Alter, Geschlecht und Schulbildung parallelisiert.

Die mit einem höheren Schizophrenierisiko behafteten biologischen Angehörigen Schizophrener wiesen - wie die schizophrenen Patienten, wenn auch in geringerem Maße - in allen BSABS-Subsyndromen eine größere Zahl von Basissymptomen auf als die Kontrollgruppe (s. Abb. 16). Die Mittelwertsdifferenzen wurden jedoch nicht nur für die Informationsverarbeitungsstörungen und die Anzeichen einer interpersonellen Verunsicherung signifikant, sondern auch für das eher unspezifische Stufe-1-Basissymptome beinhaltende Subsyndrom 'Vulnerabilität'. Letzteres wurde von den Autoren - in Übereinstimmung mit Befunden einer signifikanten Häufung affektiver Erkrankungen mit psychotischer Symptomatik, aber nicht affektiver Erkrankungen ohne psychotische Symptome bei erstgradigen Angehörigen Schizophrener (Kendler et al., 1993) - als Indiz einer generellen Disposition zur Entwicklung einer psychotischen Erkrankung interpretiert, da für diese Symptomgruppe in anderen Untersuchungen keine signifikanten Unterschiede zwischen schizophrenen und affektiven Psychosen oder gar höhere Werte bei Patienten mit affektiven Störungen gefunden wurden

(Bechdorf, Halve, Schultze-Lutter & Klosterkötter, 1998; Ebel, Gross, Klosterkötter & Huber, 1989; Klosterkötter, Gross et al., 1997).

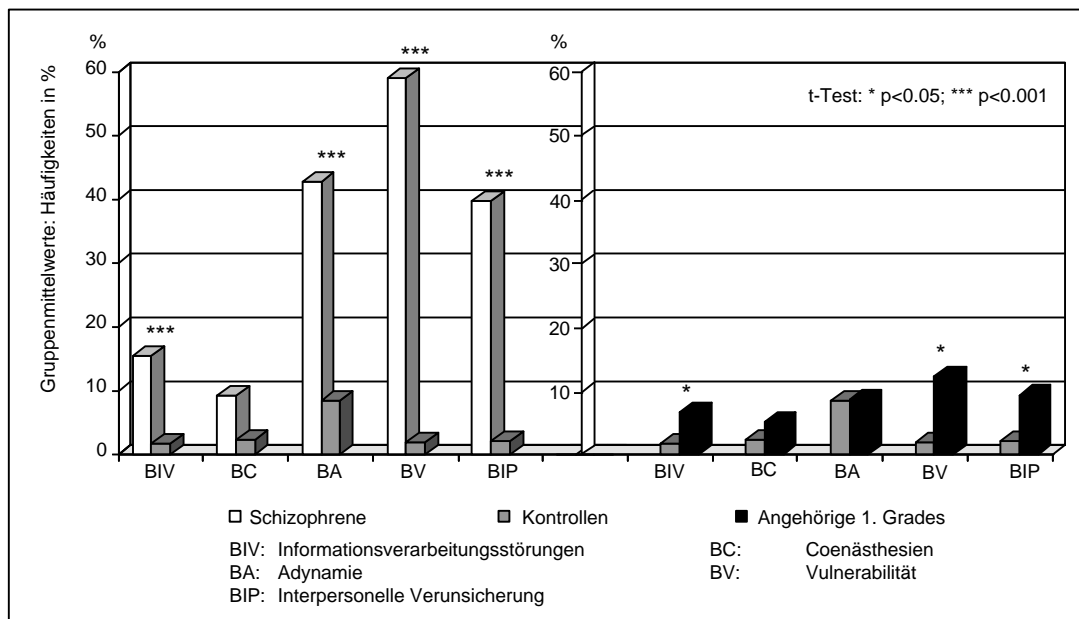


Abb. 16: Mittelwertsvergleich der BSABS-Subsyndrome über die drei Gruppen (nach Klosterkötter, Gross et al., 1997).

Aufgrund des möglicherweise genetisch fundierten phänomenologischen Überschneidungsbereichs affektiver und schizophrener Psychosen forderten Yung und McGorry (1996) vor allem bei Studien zu initialen Prodromen eine Betrachtung psychotischer Störungen als Ganzes, da die bei einer psychotischen Erstepisode vergebene Diagnose häufig, wenn Rückfälle auftreten, im weiteren Verlauf der Erkrankung revidiert werde (McGorry, 1994).

Zwei Untersuchungen sind bislang dem gezielten Vergleich von BSABS-Basissymptomen bei Personen mit schizophrenen und affektiven Störungen gewidmet worden (Bechdorf, Halve, Schultze-Lutter & Klosterkötter, 1998; Ebel, Gross, Klosterkötter & Huber, 1989). Beide Studien fanden übereinstimmend signifikant höhere Werte für Schizophrenie im Bereich interpersoneller Verunsicherung sowie höhere Werte von Patienten mit affektiven Störungen im Subsyndrom 'Vulnerabilität'. Hingegen differierten die Ergebnisse beider Studien für kognitive und coenästhetische Beschwerden sowie dynamische Defizienzen: Wurde in der älteren Arbeit (Ebel, Gross, Klosterkötter & Huber, 1989) kein Gruppenunterschied in der Häufigkeit adynamischer Beschwerdeschilderungen, aber signifikant häufiger Denk-, Wahrnehmungs- und Handlungsstörungen sowie körperliche Mißempfindungen in der schizophrenen Gruppe ermittelt, berichteten in der neueren Untersuchung depressive Patienten adynamische und coenästhetische Beschwerden signifikant zahlreicher als schizophrene Patienten; ein signifikanter Gruppenunterschied für 'Informationsverarbeitungsstörungen' wurde nicht

gefunden (Bechdorf, Halve, Schultze-Lutter & Klosterkötter, 1998). Aus den Ergebnissen letzterer Arbeit wurde geschlußfolgert, daß „die präepisodischen Störungen der Schizophrenen ... sich als *Mikroproduktivität* charakterisieren [lassen, v. Verf.], die der Depressiven als ein depressives Syndrom milder Ausprägung“ (ebd., S. 378). Ebel, Gross, Huber und Klosterkötter (1989) hingegen hatten in der Diskussion ihrer Ergebnisse besonderen Nachdruck auf den Befund häufigerer Informationsverarbeitungsstörungen bei Schizophrenen gelegt und ihn zunächst im Sinne von Berners Konzept eines Achsensyndroms endogener Psychosen (Berner, 1982) interpretiert, in dem beobachtbare Denkstörungen als Primärsymptome des endogenomorph-schizophrenen Achsensyndroms definiert sind. Später wurden die Befunde unter Bezugnahme auf Ergebnisse anderer Studien im Sinne eines Kontinuitätsmodells idiopathischer Psychosynonyme gedeutet, das die in heutigen Diagnosesystemen definierten unterschiedlichen Psychosetypen als „psychopathologische Zustands-Verlaufs-Einheiten, wobei psychopathologische Syndrome und Verlauf solcher Einheiten sich aufgrund zusätzlicher Faktoren, ..., sich [*sic*] sehr unterschiedlich gestalten können“ (Huber, 1992, S. 64) betrachtet, wobei sich schizophrene Erstrangsymptome „aus (kognitiven) Basis-symptomen, die bereits quasi „mikroproduktive“ Phänomene ... sind“ (ebd., S. 68), entwickeln.

2.1.7.3 Basissymptome als Prodromalsymptome schizophrener Erkrankungen

Doch nicht nur in retrospektiven Untersuchungen zeigten sich ‘Informationsverarbeitungsstörungen’ und ‘Interpersonelle Verunsicherung’ weitestgehend von herausragender Bedeutung für die Entwicklung einer Psychose, sondern auch in den vorläufigen Ergebnissen des prospektiven Köln-Bonn-Aachener Früherkennungsprojekts (Klosterkötter, Schultze-Lutter et al., 1997), auf dessen Daten nach Abschluß des Projekts die vorliegende Arbeit basiert und das in Kapitel 3.1.2 eingehend beschrieben ist.

Hier zeigten sich in einer Zwischenauswertung (ebd.) zum Zeitpunkt der Indexuntersuchung bei 96 Patienten mit Verdacht auf ein initiales schizophrene Prodrom signifikant höhere Mittelwerte dieser beiden Subsyndrome in der Gruppe, die im Katamnesezeitraum tatsächlich eine schizophrene Störung entwickelt hatte ($n = 56$), als in der Gruppe ohne schizophrene Entwicklung ($n = 40$; s. Abb. 17). Die Patienten hatten bei der Indexuntersuchung noch niemals positive oder negative Symptome geboten; ihre Zuordnung zu einer der beiden Gruppen (mit versus ohne Übergang in eine schizophrene Psychose) erfolgte bei der durchschnittlich acht Jahre späteren Nachuntersuchung.

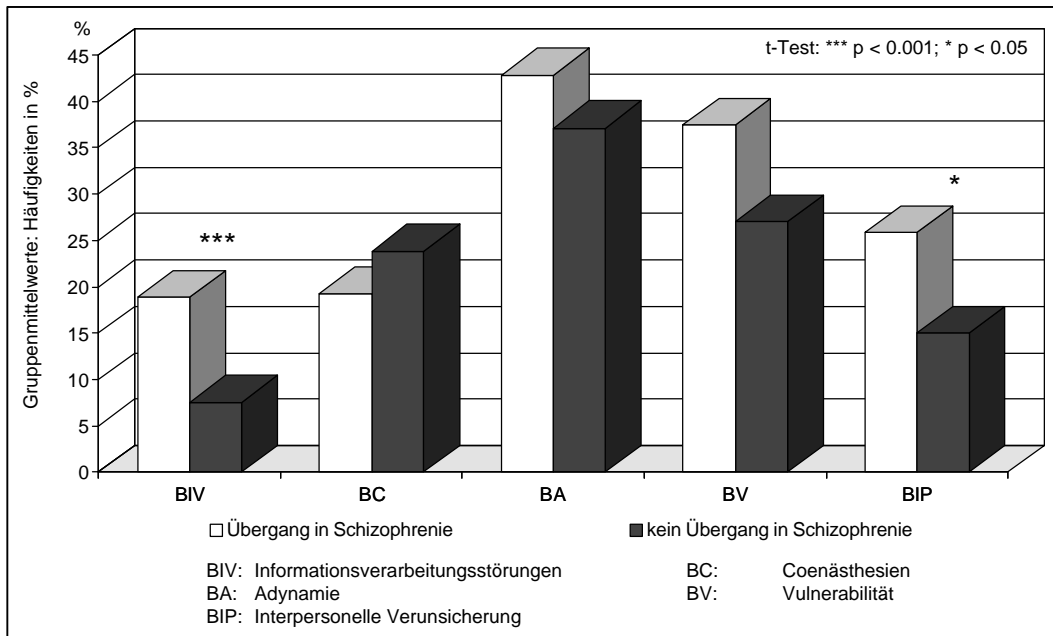


Abb. 17: Mittelwertsvergleich der BSABS-Subsyndrome über die zwei Gruppen des Köln-Bonn-Aachener Früherkennungsprojekts (nach Klosterkötter, Schultze-Lutter et al., 1997).

Kausalanalysen (LISREL 8; Jöreskog & Sörbom, 1993), die über die auf Einzelitemebene bei Personen mit einem späteren Übergang in eine schizophrene Erkrankung signifikant häufiger bei der Erstuntersuchung berichteten Basissymptome gerechnet wurden, bestätigten sowohl die Rolle der Subsyndrome 'Informationsverarbeitungsstörungen' und 'Interpersonelle Verunsicherung' als auch weitgehend deren empirische Itemgruppierung (s. Abb. 18).

Das in Abbildung 18 dargestellte Kausalmodell hatte die höchste Anpassung an die empirischen Daten und konnte 94,1% der Ursprungsvarianz erklären. Die latente Variable ξ_2 , die hierin das Subsyndrom 'Interpersonelle Verunsicherung' repräsentiert, besaß einen direkten und hoch signifikanten Einfluß auf die Entwicklung einer Schizophrenie, der einen Großteil der Varianz der Zielvariable 'Übergang in eine schizophrene Psychose' erklärte ($r^2 = 0,920$). ξ_2 wurde in diesem Modell durch die beobachteten Variablen x_6 , x_7 und x_8 beschrieben, wobei x_8 (Wahrnehmungsveränderungen am Gesicht und/oder der Gestalt anderer) clusteranalytisch dem Subsyndrom 'Informationsverarbeitungsstörungen' zugerechnet wurde, hier aber aufgrund hoher Kovarianz mit x_6 (Störung des In-Erscheinung-Tretens) und x_7 (Subjekt-Zentrismus) die latente Variable 'Interpersonelle Verunsicherung' mit repräsentiert.

ξ_1 , die latente Variable 'Informationsverarbeitungsstörungen', wie sie durch x_1 bis x_5 repräsentiert wurde, beeinflusste ξ_2 in hohem Maße, hatte aber für sich genommen nur einen geringen direkten Einfluß auf die Zielvariable. Entsprechend der dem Basissymptom-Konzept zugrunde liegenden Modellvorstellung gehen den durch Stufe-2-Basissymptomen repräsentierten latenten Variablen Stufe-1-Merkmale voraus (z_2 und z_3), und es fand sich eine gene-

relle Disposition zu schizophrenen Störungen in Form einer bei der Erstuntersuchung diagnostizierten schizotypen Persönlichkeitsstörung (z_1), deren jeweiliger kausaler Einfluß auf ξ_1 jedoch gering war ($0.231 \leq l_{j1} \leq 0.284$). Demographische Variablen, etwa Alter bei Auftreten der ersten psychischen Beschwerden, Geschlecht oder höchster erreichter Schulabschluß, hatten weder einen direkten noch einen über eine andere Modellkomponente vermittelten Einfluß auf die Zielvariable 'Übergang in eine schizophrene Psychose' (Klosterkötter, Schultze-Lutter et al., 1997).

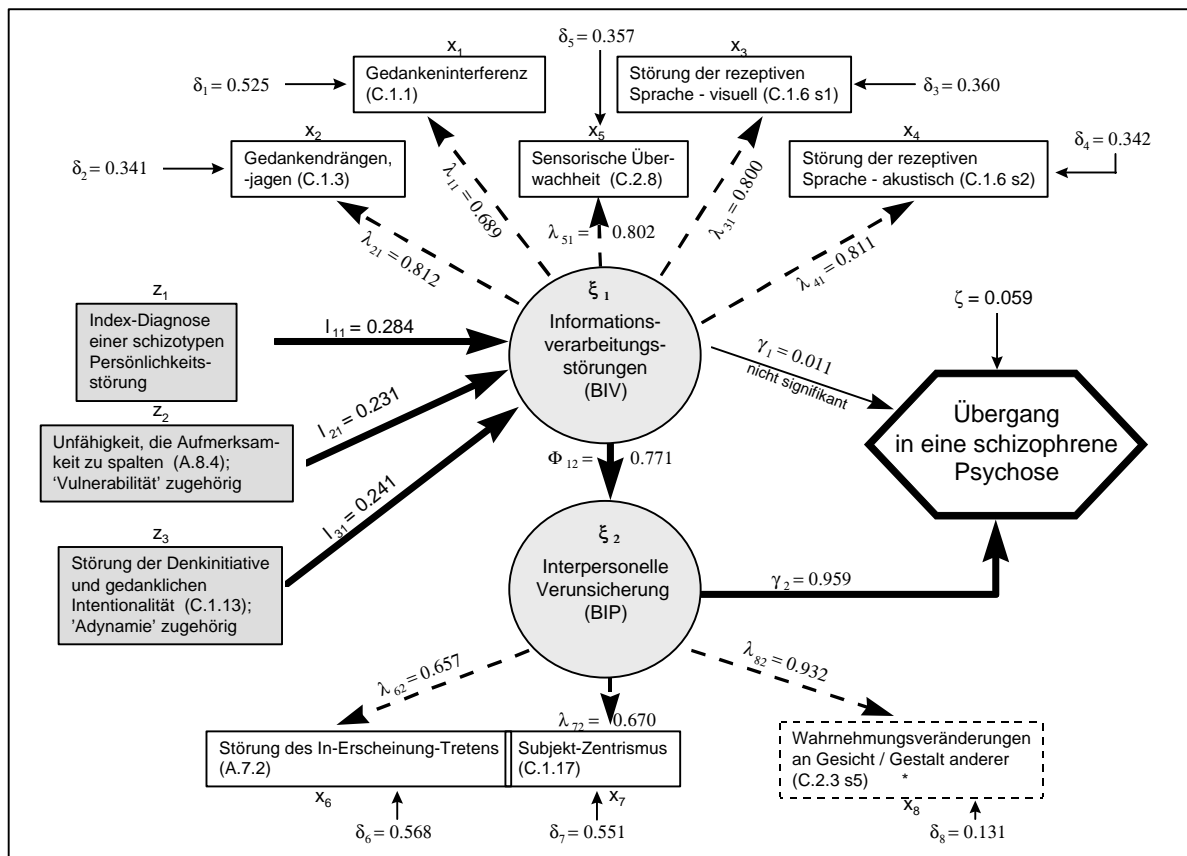


Abb. 18: Kausalmodell des Übergangs in eine schizophrene Psychose anhand der Daten des Köln-Bonn-Aachener Früherkennungsprojekts (nach Klosterkötter, Schultze-Lutter et al., 1997).

$$\chi^2 = 32,56; df = 48; p = 0,957$$

Goodness of fit-Index: 0,947; adjustierter Goodness of fit-Index: 0,914

erklärte Varianz: 0,941; Residualanteil nicht erklärter empirischer Kovarianzen: 5,2%

x_i : beobachtete Variable, die eine latente Variable repräsentiert

z_i : beobachtete Variable mit postuliertem kausalem Einfluß auf latente Variable

ξ_j : latente Variable

δ_i : Residualwert bzw. Fehlervarianz der beobachteten Variable x_i

λ_{ij} : Pfadkoeffizient, d.h. Stärke der kausalen Beziehung zwischen x_i und ξ_j

γ_j : Pfadkoeffizient zwischen ξ_j und Zielvariable

Φ_{ij} : Pfadkoeffizient zwischen latenten Variablen

l_{ij} : Pfadkoeffizient zwischen z_i und ξ_j

ζ : nicht erklärte Varianz der Zielvariable

Empirische Untersuchungen zur Eignung von Basissymptomen zur Früherkennung schizophrener Psychosen wurden somit bereits auf allen vier von Yung und McGorry (1996) beschriebenen methodischen Ebenen zur Erfassung von Prodromalsymptomatik sowie zur Entwicklung von Übergangsmo­dellen von Prodromen zu schizophrenen Psychosen (vgl. Kap. 2.1.5.3 und 2.1.5.4) durchgeführt. Ihre Ergebnisse deuten auf eine Schizophreniespezifität von Symptomen aus den Bereichen 'Informationsverarbeitungsstörungen' und 'Interpersonelle Verunsicherung'. Damit kann das Basissymptom-Konzept in seiner in der BSABS operationalisierten Form heute, nachdem sich die im DSM-III und DSM-III-R definierten Prodromalsymptome als zu unspezifisch erwiesen haben und fallengelassen wurden, als vielleicht vielversprechendste Sammlung psychometrisch erfaßbarer Prodromalsymptome einer schizophrenen Psychose betrachtet werden, weshalb seine Symptombeschreibungen auch weitgehend Eingang in die neu konzipierten Früherkennungsinstrumente der beiden weltweit führenden Forscher auf diesem Gebiet gefunden haben, der 'Scale Of Prodromal Symptoms - SOPS' von Thomas H. McGlashan von der Yale University in New Haven / USA (Miller et al., 1999) und dem 'Comprehensive Assessment of At Risk Mental States - CAARMS' von Patrick D. McGorry und Alison R. Yung aus Melbourne / Australien (Yung et al., 1998). In einer vorläufigen Auswertung der Melbourn­er Prodromstudie an Patienten mit einem möglichen Prodrom einer schizophrenen Psychose bestätigten sich die Befunde zur BSABS zum Teil erneut: Insbesondere auf psychopathologischer Ebene ließen sich signifikante Unterschiede zwischen in eine akute Psychose übergegangenen und (noch) nicht übergegangenen Patienten aufdecken, und unter diesen Symptomen fanden sich auch eine Reihe aus der BSABS übernommener Basissymptome (McGorry et al., 1999).

Um jedoch ein praktikables Früherkennungsinstrument zu erstellen, ist es notwendig eine kleine, aber dennoch hoch prädiktive Anzahl von Symptomen zu finden, die im Individualfall eine möglichst eindeutige und verläßliche Unterscheidung zwischen einem schizophrenen Prodrom und anderen psychischen Erkrankungen oder Fluktuationen des psychischen Befindens treffen kann und auch der Tatsache gerecht wird, daß in der alltäglichen ärztlichen Praxis nur eine begrenzte Zeitspanne für die Untersuchung eines einzelnen Patienten zur Verfügung steht (Jackson et al., 1996). Für die Auswahl solcher sensitiver, spezifischer und prädiktiver Symptome aus einer größeren Menge potentieller prädiktiver Variablen, die auf psychometrischer Ebene meist nur ein nominales Meßniveau erreichen, und die Generierung eines hierauf beruhenden validen und reliablen Prognosemodells wurden gerade in den letzten Jahren eine Reihe ganz unterschiedlicher Techniken und Strategien entwickelt (Agresti, 1996; Andreß, Hagen­aars & Kühnel, 1997; Ashby et al., 1992; Marder et al., 1991; Marshall, 1986, 1995, 1997; Marshall, Grover, Henderson & Hammermeister, 1994; Pawlak, 1984; Pawlak, Wong & Ziarko, 1988), die bislang aber kaum im Hinblick auf die für die Ent-

wicklung eines ökonomischen und in der Einzelfalldiagnostik einsetzbaren Früherkennungsinstruments entscheidende Frage, die Vorhersageleistung ihrer generierten Modelle und damit auf die Verlässlichkeit ihrer Prognose, verglichen wurden.

2.2 MATHEMATISCHE METHODEN ZUR GENERIERUNG VON VORHERSAGEMODELLEN

In der Medizin stellt die Prognose traditionell seit mehr als 2000 Jahren die Vorhersage über den Krankheitsverlauf eines einzelnen Patienten dar, die der Arzt aufgrund seiner Erfahrungen mit vorangegangenen Patienten trifft. In diese prognostische Entscheidungsfindung wird eine Anzahl von Merkmalen, die die Ausgangsverfassung früherer Patienten charakterisierten, sowie deren klinischer Verlauf einbezogen. Hierbei werden nur diejenigen Merkmale bei der Prognose berücksichtigt, die in der Vergangenheit als Vorboten eines bestimmten Krankheitsverlaufs konsistent aufgetreten waren. Diese Merkmale werden bei einem neuen Patienten auf ihr Vorliegen hin überprüft und bilden damit die Grundlage der Verlaufsprognose. Bei dieser traditionellen Form klinischer Entscheidungsfindung, deren Basisstrategie einen 'selektiven Ausschluß' (engl.: „selective exclusion“; Koss & Feinstein, 1971, p. 448) darstellt, werden von der möglichen Vielzahl der Beobachtungen über einen Patienten alle diejenigen außer Acht gelassen, die nicht maßgeblich für die Vorhersage des Krankheitsverlaufs oder auch die Diagnose der Erkrankung erscheinen, beispielsweise das Vorhandensein einer Erkältung bei der Beurteilung einer möglicherweise beginnenden Schizophrenie. Diese Vorhersagestrategie hat sich in der langen Geschichte der Medizin sowohl aus logischen wie aus mnestischen Gründen als sinnvoll erwiesen: Aus logischer Sicht bietet sie eine Möglichkeit, sinnvolle Zusammenhänge in der unübersichtlichen Komplexität von Daten vorangegangener Patienten aufzudecken; aus mnestischer Sicht hilft sie, diese Datenmenge auf ein kognitionspsychologisch erinnerbares Maß zu reduzieren (ebd.).

Die fortschreitende Computerisierung kann heute jedoch helfen, die Grenzen menschlichen Erinnerungsvermögens zu überwinden; selbst privaten Praxen stehen mit den modernen hochleistungsfähigen Personal Computern die technischen Möglichkeiten zur Verarbeitung auch sehr großer Datenmengen zur Verfügung, so daß für den Mediziner nicht mehr die Notwendigkeit besteht, die Details und Verläufe einzelner Patienten möglichst genau im Gedächtnis zu behalten. Da aber irrelevante, nicht mit der zu beurteilenden Erkrankung zusammenhängende Informationen zu einer verzerrten oder falschen Prognose führen können, besteht das Problem fort, eine adäquate, logische Methode zur Informationsverarbeitung

heranzuziehen, die nicht nur ein valides und reliables Vorhersagemodell generiert, sondern auch die hierfür relevanten Patientendaten auswählt (ebd.).

Waren aus diesem Grund empirische Vorhersagemodelle für die Prognose des Krankheitsverlaufs schon lange von Interesse, so kommt ihnen im Zuge des Paradigmenwandels vom klassischen kurativen Ansatz der Krankheitsbekämpfung hin zu einem präventiven Ansatz der Krankheitsvermeidung in der gesamten medizinischen Forschung ein immer größeres Gewicht zu. Die psychiatrische Forschung schloß sich dieser Entwicklung hin einer zumindest sekundären Prävention psychischer Erkrankungen allerdings erst in neuerer Zeit an, was sich unter anderem dadurch erklärt, daß die auf diesem Gebiet erhobenen Daten, insbesondere auf psychopathologischer Ebene, meist ein nur maximal ordinales Skalenniveau erreichen, in einer Vielzahl der Fälle aber kategorial bzw. dichotom sind. Da die meisten der konventionellen statistischen Methoden zur Generierung von Vorhersagemodellen, wie etwa diskriminanz- oder regressionsanalytische Verfahren, jedoch lange Zeit intervallskalierte Variablen voraussetzten, standen bei Nominaldaten historisch gesehen zunächst χ^2 -Methoden zur Analyse von Kontingenztafeln im Mittelpunkt. Viele der heute verfügbaren Verfahren zur kategorialen Datenanalyse waren noch vor 15 Jahren entweder nicht existent, kaum bekannt oder für die breite Masse der Forscher nicht anwendbar, da die Computerentwicklung hinter der methodischen Entwicklung zurücklag (Preisser & Koch, 1997).

2.2.1 Logistische Regressionsanalyse

Zur Generierung von Vorhersagemodellen bei Daten mit Nominalskalenniveau sind bis heute Logitmodelle, basierend auf den bereits 1958 von Cox entwickelten logistischen Regressionsverfahren, die am häufigsten verwandten Methoden, um simultan die Beziehungen von mehreren unabhängigen Variablen zu einer dichotomen abhängigen Variablen zu erheben (Ashby et al., 1992; Harrell, Lee, Matchar & Reichert, 1985). Mit diesen Verfahren können Zusammenhänge erkannt und erklärt und die Wahrscheinlichkeit der abhängigen Variablen in Relation zu einer Anzahl prädiktiver unabhängiger Variablen geschätzt bzw. prognostiziert werden.

Das multiple logistische Regressionsmodell hat dabei die Form folgender Gleichung:

$$\log (p/[1 - p]) = \beta_0 + \beta_1x_1 + \dots + \beta_ix_i$$

- wobei: p: die Wahrscheinlichkeit der abhängigen binären Variable
x_i: die potentiell prädiktive unabhängige Variable, Regressor
β_i: Maß des Einfluß' der unabhängigen Variablen x_i, Regressionskoeffizient
β₀: konstantes Glied, das die Basiswahrscheinlichkeit nach Auswahl der Regressoren repräsentiert

Um die Wahrscheinlichkeiten der alternativen Ausprägungen der abhängigen Variablen y , die Werte im Bereich 0 bis 1 darstellen, zum linearen Prädiktor ' $\beta_0 + \beta_1 x_1 + \dots + \beta_i x_i$ ', der prinzipiell beliebig große und auch negative Werte ($-\infty$ bis ∞) annehmen kann, in Relation zu setzen, wird die Transformation des linearen Prädiktors über den natürlichen Logarithmus des Verhältnisses der Wahrscheinlichkeiten beider Ausprägungen der abhängigen Variable (engl.: „odds“) vorgenommen, $\log(p/[1 - p])$, wobei $p = p(y = 1)$ und $1 - p = p(y = 0)$. Dadurch werden die sogenannten *log odds* oder auch *logits* ermittelt, denen diese Form der Regressionsanalyse seinen Namen verdankt.

Nach der Bestimmung eines a priori festgelegten Schwellenwerts für die Basiswahrscheinlichkeit im Bereich 0 bis 1 wird durch Berechnung der Regressionskoeffizienten der ermittelten Regressoren eine Klassifikationsregel erstellt, anhand derer die Objekte einem der beiden Ausgänge zugeordnet werden können, je nachdem, ob die für sie ermittelte Wahrscheinlichkeit unter- oder oberhalb des Schwellenwerts liegt. Im Gegensatz zu Korrelationsanalysen zur Ermittlung der Größe eines irgendwie gearteten potentiellen Zusammenhangs wird hierbei bereits im Vorfeld eine eindeutige Richtung der Beziehung zwischen unabhängigen und abhängiger Variable implizit angenommen, also „Je-Desto-Beziehungen“ (Backhaus, Erichson, Plinke, Schuchard-Fischer & Weiber, 1987, S. 3) zwischen den Merkmalen und dem Zielkriterium unterstellt.

Ist das Regressionsmodell in seiner einfachsten, grundlegenden Form formuliert, impliziert diese Additivitätsannahme, daß die unabhängigen Variablen mit dem Zielkriterium in additiver Weise verknüpft sind. Bei binären 0/1-kodierten unabhängigen Variablen hat dies eine reine Summierung der Gewichte der vorliegenden 1-kodierten Variablen für die Berechnung der Wahrscheinlichkeit der abhängigen Variable zur Folge. Im Falle der Früherkennung von Schizophrenie bedeutet dies, daß dem logistischen Modell die Annahme immanent ist, daß es eine adäquate Abbildung der realen Verhältnisse ist, daß die Erkrankungswahrscheinlichkeit mit einer steigender Zahl von potentiell prädiktiven Symptomen zunimmt und andere Beziehungen zwischen den Symptomen nicht bestehen. Potentielle nicht-lineare Zusammenhänge zwischen den Variablen bleiben deshalb bei der Analyse unberücksichtigt. Damit ist die Additivitätsannahme eines der am häufigsten gegen die Verwendung von Regressions-, aber auch von Diskriminanzanalysen angeführten Argumente. So sehen etwa Harrell und Kollegen (1985) gerade mit der Entwicklung immer leistungsfähigerer Computerprogramme die Gefahr einer Anwendung multipler logistischer Regressionsanalysen ohne vorherige Prüfung der inhärenten Modellannahmen, wobei Annahmen bezüglich Linearität und Verteilung der abhängigen Variable bei logistischen Regressionsanalysen keine oder zumindest nur eine untergeordnete Rolle spielen. Der Additivitätsannahme sollte hingegen - soweit sie nicht hinreichend empirisch oder theoretisch fundiert ist - durch die Berücksichtigung von In-

teraktionen der Prädiktorvariablen in Form von Kreuz-Produkt-Termen bei der Modellkalkulation Rechnung getragen werden. Ein Nachteil der logistischen Regressionsanalyse ist es aber, daß sie a priori keine rechnerische Grundlage zur Aufdeckung möglicherweise relevanter Interaktionen der unabhängigen Variablen bietet und die Berücksichtigung aller möglichen Interaktionen zu einer entsprechenden Erhöhung der benötigten Stichprobengröße führt, da jeder Kreuz-Produkt-Term seinerseits als zusätzliche unabhängige Variable fungiert. Mögliche relevante Kreuz-Produkt-Terme müssen deshalb entweder aufgrund theoretischer Vorüberlegungen eingeführt oder mit Hilfe anderer mathematischer Verfahren wie etwa artifiziellen neuronalen Netzwerken, Grob-Mengen-Modellen oder Partitionierungsverfahren aufgedeckt werden (Lette et al., 1994). Die Einführung von Kreuz-Produkt-Termen auf der Basis theoretischer Vorüberlegungen erfordert jedoch eine gute Kenntnis sowohl des untersuchten Bereichs als auch der verwandten Methode, da es sonst leicht zu Fehlinterpretationen von Daten kommen kann, insbesondere wenn Datenerhebung und -auswertung, wie in der medizinischen Forschung vielfach üblich, in unterschiedlichen Händen liegen (Appleton, 1995).

The degree of understanding which the statistician has for the problem which the clinician is trying to answer, and the understanding which the clinician has of the capabilities, though not necessary details, of statistical methodology are important to the success of any collaborative project. The strength and the weakness of methods of statistical analysis is that they make simplifying assumptions. ... Sometimes such assumptions [etwa Verteilungs-, Linearitäts- und/oder Additivitätsannahmen; Anm. v. Verf.] are perfectly reasonable, the methods they unleash provide good summaries of the data, and inferences which result give genuine insight into the clinician's problem.

Sometimes it is possible that both statistician and clinician forget that the assumptions have been made, or make assumptions that are not justified by real-life aspects of the system they are modelling, or they fail to investigate how dependent their inferences are on the model which they have imposed on their data. (ebd., p. 185)

Ist bei der Versuchsplanung und Datenerhebung zudem noch nicht klar, welche unabhängigen Variablen im Modell zu berücksichtigen sind, etwa weil gemäß den Forderungen von Jackson und Kollegen (1996) möglichst viele potentielle Prodromalsymptome erhoben wurden, so können mit Hilfe der schrittweisen logistischen Regression nicht nur die Modellgleichung, sondern darüber hinaus auch die in die Gleichung eingehenden 'wichtigen' Variablen ermittelt werden. Hierbei wird bei jedem Selektionsschritt über χ^2 -Tests die Güte der Modellanpassung einerseits und der einzelnen potentiellen Einflußfaktoren x_i andererseits bestimmt. So wird etwa über Wald- χ^2 -Tests zu den einzelnen Regressionskoeffizienten getestet, ob der jeweilige Faktor, also die jeweilige unabhängige Variable oder der jeweilige

Kreuz-Produkt-Term, nach Adjustierung um alle anderen Parameter von signifikantem Einfluß unter der Testhypothese $H_0: \beta_i = 0$ ist.

Bei dem Versuch, über ein solches schrittweises Verfahren zu einem möglichst einfachen, ökonomischen Logitmodell zu gelangen, werden jedoch nach Ansicht von Harrell und Kollegen (1985) sehr häufig Vorhersagemodelle anhand einer zu hohen Anzahl von unabhängigen Variablen bzw. zu geringer Stichprobengröße generiert, die eine Überanpassung („overfit“; ebd., p. 1071) an die empirischen Daten aufweisen und in der Folge der Validierung an einer unabhängigen Stichprobe nicht standhalten. Dies gilt insbesondere dann, wenn die Zahl der Personen mit dem Zielkriterium, etwa dem Übergang in eine Schizophrenie, in der Stichprobe gering ist. Für eine schrittweise Variablenselektion wird daher von ihnen als grober Richtwert ein Verhältnis von 1:10 zwischen der Zahl der in die Analyse eingehenden Regressoren und der Zahl der in der Stichprobe enthaltenen Personen mit dem Zielkriterium angegeben (Harrell, Lee, Califf, Pryor & Rosati, 1984). Diese Richtlinie ist jedoch sehr konservativ und wesentlich restriktiver als das von Backhaus und Koautoren (1987) empfohlene Verhältnis von 1:2.

Darüber hinaus konnte für logistische Regressionsanalysen gezeigt werden, daß sich schon bei relativ geringfügigen Änderungen der Wahl der in die Analyse eingehenden unabhängigen Variablen die in den Vorhersagemodellen enthaltenen Regressoren erheblich unterscheiden (Harrell et al., 1984). Somit stellt sich gerade in der klinischen Forschung, in der Stichproben meist nicht frei verfügbar sind, unabhängig von den unterschiedlichen Empfehlungen für das Verhältnis von Stichprobengröße und Zahl der unabhängigen Variablen das Problem, welche der potentiell einflußreichen Variablen in der Analyse berücksichtigt werden sollen. Zur Lösung dieses Problems werden von Harrell und Mitarbeitern (1985) drei alternative Vorgehensweisen vorgeschlagen:

1. vorangestellte Variablenselektion nach statistischer Signifikanz,
2. vorherige Hauptkomponentenanalyse und anschließende Nutzung der Faktoren als unabhängige Variablen oder
3. vorangehende Clusteranalyse der unabhängigen Variablen mit anschließender Verbesserung der Cluster anhand des klinischen Urteils und der Cluster-Variablen-Korrelationsmatrix und nachfolgender Verwendung der Cluster als unabhängige Variablen.

Ein Vergleich dieser drei Vorgehensweisen zeigte in der Studie der Autoren eine Überlegenheit des Vorhersagemodells nach Variablenreduktion durch eine Hauptkomponentenanalyse in einer Validierungsstichprobe bei nahezu gleicher Güte des Modells über die Cluster in größeren Stichproben (ebd.). Eine Überlegenheit der schrittweisen logistischen Regressi-

onsanalyse von Clustern im Vergleich zu der von Hauptkomponenten oder Einzelvariablen fanden hingegen Marshall, Grover, Henderson und Hammermeister (1994).

Ein weiteres Problem, das zu einer mangelhaften Replikation des Vorhersagemodells an einer Validierungsstichprobe führen kann und dem Problem der Fehleradjustierung bei multiplen Paarvergleichsstudien entspricht, ergibt sich aus der Vielzahl von Vergleichen bei einer großen Anzahl von unabhängigen Variablen. Dabei müssen bei einem dichotomen Ausgang und k Regressoren $[k(k - 1)]/2$ Paarvergleiche vorgenommen werden, so daß die Möglichkeit, daß einige der unabhängigen Variablen zufällig signifikant werden und damit Aufnahme in das Modell finden, bei $1 - (1 - \alpha)^{[k(k - 1)]/2}$ liegt. Zudem besteht die Gefahr, daß die Wichtigkeit einiger Variablen aufgrund von Idiosynkrasien in der Berechnungsstichprobe überschätzt wird und so ein an die Berechnungsstichprobe überangepaßtes Modell berechnet wird, das die tatsächlichen Zusammenhänge zwischen abhängiger und unabhängigen Variablen nur ungenügend bzw. verzerrt wiedergibt (Harrell et al., 1985).

Um diesen Problemen schon bei der Versuchsplanung entgegenzutreten, sollte zum einen das Studiendesign dem späteren Vorhersagemodell angepaßt, d.h. prospektiv, und zum anderen eine Validierungsstichprobe vorhanden sein, um die tatsächliche Klassifizierungsleistung des Vorhersagemodells realistischer abschätzen zu können. So weist eine größere Diskrepanz zwischen der Vorhersageleistung in der Berechnungs- und Validierungstichprobe auf eine Überanpassung des Modells an die Daten der Berechnungsstichprobe hin (ebd.).

Aufgrund ihrer Überlegungen kamen Harrell und Kollegen (1985) im Hinblick auf die Anwendbarkeit von Regressionsanalysen vor allem im Vergleich zu alternativen Stratifizierungsverfahren zu dem folgenden Schluß:

When a regression analysis is not performed carefully, other methods such as recursive partitioning can be more accurate. When the analysis is done carefully, however, regression models can make prognostic estimates that are often as good and sometimes better than those obtained using stratification methods. (ebd., p. 1077)

2.2.2 Nonparametrische Diskriminanzanalyse

Regressions- bzw. Logitmodelle werden meist den 'älteren' diskriminanzanalytischen Verfahren zur Analyse binärer Daten vorgezogen, da angenommen wird, „that they will not perform better than logistic regression due to the high degree of non-normality of the data“ (Marshall et al., 1994, p. 1502). Trotz formaler Ähnlichkeiten ihres linearen Prädiktors bestehen gravierende modelltheoretische Unterschiede zwischen Regressions- und Diskriminanzanalysen: Während bei der Regressionsanalyse die abhängige Variable als Zufallsvariable behandelt

wird und die unabhängigen Variablen als fix angesehen werden, verhält es sich bei der Diskriminanzanalyse genau umgekehrt, die abhängigen Variablen werden als fix und die unabhängigen Variablen als zufällig variiert behandelt. Auf dieser Grundlage wird im Gegensatz zu Regressionsmodellen bei diskriminanzanalytischen Modellen eine Diskriminanzfunktion berechnet, deren einzelne, auf ihr liegende Diskriminanzwerte L die prädiktiven Informationen aller enthaltenen unabhängigen Variablen im Individualfall zusammenfassen:

$$L = b_0 + b_1x_1 + \dots + b_ix_i$$

wobei: L: der Diskriminanzwert der abhängigen binären Variable
 x_i : die potentiell prädiktive unabhängige Variable
 b_i : Maß des Einfluß' der unabhängigen Variablen x_i , Diskriminanzkoeffizient
 b_0 : konstantes Glied

Während bei kontinuierlichen Variablen unter der Annahme einer Normalverteilung der Variablen die Diskriminanzfunktion anhand generalisierter quadrierter Distanzen berechnet wird, erfolgt die Berechnung bei nicht metrisch skalierten Variablen über nonparametrische Schätzungen gruppenspezifischer Wahrscheinlichkeitsdichten (engl.: „probability densities“). Hierbei stehen zwei Methoden zur Verfügung, kernel und k-nearest-neighbour, um nonparametrische Schätzungen der Dichte für jede Gruppe zu berechnen und darüber ein Klassifikationskriterium zu generieren. Die Proximität wird hierbei anhand quadrierter Euklidischer oder Mahalanobis Distanzen, D^2 , berechnet. Aufgrund der so generierten Klassifikationsfunktion wird ein Objekt i in diejenige Gruppe g eingeordnet, der es am nächsten liegt, d.h. zu deren Centroid, der aus den Daten geschätzten gruppenspezifischen Dichte, seine Distanz minimal wird:

$$D^2_{ig} = \sum (L_{ki} - \bar{L}_{kg})^2$$

wobei: \underline{L}_{ki} : der Diskriminanzwert von Objekt i bzgl. Diskriminanzfunktion k
 \bar{L}_{kg} : Centroid von Gruppe g bzgl. Diskriminanzfunktion k

Während die schrittweise logistische Regressionsanalyse die Möglichkeit zur gleichzeitigen Modellgenerierung und Variablenselektion bietet, muß dies bei der Diskriminanzanalyse sukzessiv erfolgen, indem zunächst eine schrittweise Diskriminanzanalyse durchgeführt und anschließend eine nonparametrische Modellfunktion über die selektierten unabhängigen Variablen ermittelt wird. Eine spezielle schrittweise Diskriminanzanalyse für nonparametrische Daten steht bislang noch nicht zur Verfügung, doch kann nach den Anwendungsempfehlungen des SAS (SAS Institute Inc., 1990) die für metrische, normalverteilte Daten bestehende schrittweise Analyse deskriptiv auch auf nonparametrische Daten angewandt werden, wenn ihr eine dem Datenniveau entsprechende Diskriminanzanalyse folgt. Wie bei der schrittweisen logistischen Regressionsanalyse wird auch bei der schrittweisen Diskriminanzanalyse

pro Selektionsschritt diejenige unabhängige Variable ausgewählt, die ein bestimmtes Gütemaß maximiert bzw. minimiert, wenn Wilks' λ als inverses Gütemaß verwandt wird.

Die unterschiedliche Gewichtung von Merkmalen durch Koeffizienten, wie sie sowohl bei Regressions- als auch bei Diskriminanzanalysen erfolgt, und ihre Aufsummierung zu einem individuellen Referenzwert, der die Basis der individuellen Prognose darstellt, entspricht nach Koss und Feinstein (1971) einer Entscheidungsstrategie des 'abgestuften Einschlußes' („graded inclusion“; ebd., p. 449). Damit entsprechen diese statistischen Vorhersagestrategien zumindest in ihrer nicht-schrittweisen Form nicht der traditionellen medizinischen Entscheidungsfindung durch selektiven Ausschluß, die durch Eliminierung unwichtiger Information zur Bildung von Subgruppen unterschiedlicher Merkmalskombinationen und zur individuellen Vorhersage aufgrund der Charakteristiken des mit der adäquaten Subgruppe assoziierten Krankheitsverlaufs führt. Die prognostische Information über solche durch eine begrenzte Anzahl bestimmter Merkmalsausprägungen definierten Subgruppen mit unterschiedlichem Krankheitsverlauf steht jedoch auch bei den mit schrittweisen multivariaten Regressions- und Diskriminanzanalysen erzielten Vorhersagemodellen dem Kliniker nicht zur Verfügung, da diese nicht in der Lage sind, etwaige Subgruppen mit möglicherweise unterschiedlichen relevanten Prädiktoren zu identifizieren. Koss und Feinstein schlugen deshalb schon 1971 die Methode der rekursiven Partitionierung als eine alternative Analysemethode vor, die aufgrund ihrer sequentiellen, eher der klinischen Logik als mathematisch-theoretischen Kalkulationen folgenden Strategie der konventionellen klinischen Entscheidungsfindung näher stehe, für den Praktiker in seiner Entscheidungsfindung transparenter sei und keine nebulösen Generalisierungen über unspezifizierte klinische Entscheidungen enthalte.

2.2.3 Rekursive Partitionierung

Vor allem im Zuge des seit einigen Jahren wachsenden Interesses an Computerunterstützter Diagnostik, besonders in Bereichen, in denen die Diagnose von einer Vielzahl von Faktoren abhängt, deren Spezifität noch unklar ist, wird heute vermehrt versucht, Programme zu entwickeln, die mit Hilfe von sowohl statistischen als auch von 'Expertensystem'-Ansätzen die Diagnosefindung vereinfachen. Ein solcher integrativer Ansatz ist die sogenannte *rekursive* oder *hierarchische Partitionierung*, ein der nonparametrischen Diskriminanzanalyse verwandtes Verfahren, das die Daten basierend auf Merkmalen, die den gewünschten Endpunkt vorhersagen, in immer kleinere Untergruppen stratifiziert (Cook & Goldman, 1984).

Ursprünglich wurde die Methode der rekursiven Partitionierung für die 'Regression' mit einer

intervallskalierten abhängigen Variable entwickelt. Sie kann jedoch auch auf eine nominalskalierte kategoriale abhängige Variable zur Bildung von Klassifikationsregeln angewandt werden. Eine Reihe von Verfahren stehen inzwischen für die Berechnung rekursiver Partitionierungsmodelle oder verwandter Modelle zur Verfügung (AID: Morgan & Sonquist, 1963; C 4.5: Quinlan, 1993; CART: Breiman, Friedman, Olshen & Stone, 1984; CHAID: Kass, 1980; CLS: Quinlan, 1983; FACT: Loh & Vanichsetakul, 1988; HICLAS: DeBoeck & Rosenberg, 1988; KnowledgeSeeker: Biggs, De Ville & Suen, 1991; MARS: Friedmann & Roosen, 1995; S routines: Clark & Pregibon, 1992; TREEMIN: Forsyth, Clarke & Wright, 1994). Obwohl diese Methode zunächst für den Einsatz im sozialwissenschaftlichen Bereich konzipiert wurde, fand sie rasch Anwendung in der medizinischen und auch vereinzelt in der psychiatrischen Forschung (Fleiss, Spitzer, Cohen & Endicott, 1977; Spitzer & Endicott, 1968, 1969; Wing, 1970) und ist Grundlage eines kürzlich erstellten psychiatrischen Expertenprogramms zur Diagnosestellung nach ICD-10, DSM-IV und DSM-III-R, dem 'Monash Interview for Liaison Psychiatry - MILP' (Yap & Clarke, 1996).

Während logistische Regressionsmodelle die Wahrscheinlichkeit einer Ausprägung der abhängigen Variablen in Abhängigkeit von den Ausprägungen der unabhängigen Variablen beschreiben und die Diskriminanzanalyse eine einzige Wertefunktion bestimmt, die die prädiktive Information aller in ihr enthaltenen unabhängigen Variablen zusammenfaßt, verzichtet die rekursive Partitionierung auf die Spezifizierung eines immanenten parametrischen Modells und unterteilt die Daten basierend auf den Ausprägungen der unabhängigen Variablen in immer kleinere Gruppen, die die Ausprägungen der abhängigen Variable beschreiben (Cook & Goldman, 1984). Die Ausprägungen der abhängigen Variablen stellen hierbei die gewünschten Endpunkte des hierbei entstehenden Entscheidungsbaums oder Dendrogramms (s. Abb. 19) dar. Insofern steht das Ergebnis einer hierarchischen Partitionierung dem klinischen Alltag näher als etwa die bei Regressions- oder Diskriminanzanalysen gewonnenen Symptomgewichtungen, da für die diagnostische Entscheidung meist das Vorliegen bzw. Nicht-Vorliegen von Symptomen maßgeblich ist (Marshall, 1986).

Die Grundzüge der rekursiven bzw. hierarchischen Partitionierung seien im folgenden erläutert: Gegeben sei die abhängige Variable y mit J nominalen Klassen, y_0, \dots, y_{J-1} , und eine Menge von p unabhängigen Variablen x bzw. Indikatoren von y , $\mathbf{x} = (x_1, \dots, x_p)$. Die Menge möglicher Werte, die \mathbf{x} annehmen kann, sei X . Beispielsweise sei y_j einer von J möglichen diagnostischen oder prognostischen Ausgängen und \mathbf{x} eine Menge von Symptomen. Die rekursive Partitionierung bildet aus den Rohdaten zu y und \mathbf{x} eine Aufteilung von X in J Untermengen A_0, \dots, A_{J-1} dergestalt, daß jede Untermenge A_j hinsichtlich y_j 'rein' ist, d.h. nur eine mögliche Ausgangsalternative y_j enthält. Eine Beobachtung mit bekanntem \mathbf{x} und unbekanntem y kann der Klasse y_j dann zugeordnet werden, wenn \mathbf{x} zu A_j gehört. Eine solche

Klassifizierungen bzw. Vorhersagen erst erlaubende Partition entsteht durch die wiederholte Aufteilung der Stichprobe in Untergruppen mit zunehmender *Reinheit* („purity“, Marshall, 1986, p. 518) hinsichtlich y entsprechend der Werte der Elemente von \mathbf{x} . Die Struktur einer solchen Aufteilung läßt sich durch ein Dendrogramm repräsentieren, in dem sich die Äste, d.h. die Untergruppen, aufteilen, um neue Äste bzw. neue, kleinere Untergruppen zu bilden (s. Abb. 19). In einem solchen Dendrogramm ist jede der gebildeten Untergruppen, Klasse y_j Untergruppe, rein hinsichtlich einer der Klassen der abhängigen Variablen y_j . Die Attribute bzw. Variablenausprägungen, die alle Klasse y_j Untergruppen definieren und von denen theoretisch jede beliebige Anzahl vorliegen kann, bilden zusammengenommen A_j . Für die Bestimmung der ‘besten’ Größe eines Dendrogramms wurden verschiedene Kriterien vorgeschlagen, von denen Marshall (1986) die im ‘Classification and Regression Tree System - CART’ (Breiman et al., 1984) implementierte Methode einer Straffunktion für die Komplexität des Dendrogramms zur Relativierung der steigenden Reinheit der Untergruppen bei gleichzeitiger Verringerung ihrer Stichprobengröße als besten Lösungsansatz für dieses Problem betrachtet.

Bei dem Spezialfall einer dichotomen abhängigen Variablen mit $J = 2$ Ausprägungen muß eine Teilung von X in zwei Regionen A_0 und A_1 erfolgen. Hierbei ist A_1 genau die komplementäre Menge von A_0 , so daß durch die alleinige Spezifizierung von A_0 die gesamte Partition bereits exakt festgelegt ist, deshalb wird in diesem Fall A_0 als die Teilung von X bezeichnet. In der Medizin, insbesondere in der Psychiatrie, werden zudem die unabhängigen Variablen x_i häufig ebenfalls binäre Variablen sein, beispielsweise psychopathologische Symptome, die lediglich auf ihr (Nicht-)Vorhandensein hin überprüft werden. Damit induziert auch jedes x_i eine dichotome Partition, indem es etwa die Werte 0 (Symptom nicht vorhanden) und 1 (Symptom vorhanden) annehmen kann. Zur Vereinfachung soll im folgenden die Notation X_i die Region von X bezeichnen, in der x_i den Wert 0 annimmt, $X_i = \{\mathbf{x}; x_i = 0\}$. Eine mögliche Partition stellt somit $A_0 = X_i$ dar; die komplementäre Region mit $x_i = 1$ ist $A_1 = {}^kX_i$.

Kombinationen verschiedener Attribute definieren komplexere Partitionen von X . Sind etwa die binären unabhängigen Variablen x_i (Depersonalisation) und x_j (Gedankenblockierung) gegeben, stellt die Attributkombination keine Depersonalisation und keine Gedankenblockierung eine mögliche Partition von X dar, $A_0 = X_i \cap X_j$ wobei $X_i = \{\mathbf{x}; x_i = 0\}$ und $X_j = \{\mathbf{x}; x_j = 0\}$. Die entsprechende komplementäre Menge zu A_0 ist $A_1 = {}^kX_i \cup {}^kX_j$ wobei ${}^kX_i = \{\mathbf{x}; x_i = 1\}$ und ${}^kX_j = \{\mathbf{x}; x_j = 1\}$, Depersonalisationserleben oder Gedankenblockierung ist vorhanden. Da hier das oder kein ausschließendes ist, können auch beide Phänomene vorhanden sein.

Es konnte gezeigt werden, daß durch die Kombination von binären Attributen $2^{(2 \text{ hoch } p) - 1} - 1$ mögliche Pfade bei der Bildung einer dichotomen Partitionierung gegeben sind (Marshall,

1986). Jede Aufteilung kann hierbei durch eine Boolesche Funktion von X_1, \dots, X_p mit den Verbindungen \cap ('geschnitten') und \cup ('vereinigt') sowie dem komplementären Operator $\bar{}$ ausgedrückt werden. Eine direkte Suche nach der besten Aufteilung in den Daten wird damit schnell und durchführbar; so ergeben sich etwa bei fünf in die Berechnung eingehenden unabhängigen Variablen bereits $2^5 - 1 = 247$ mögliche Pfade. Hierfür bietet die rekursive Partitionierung eine alternative Strategie.

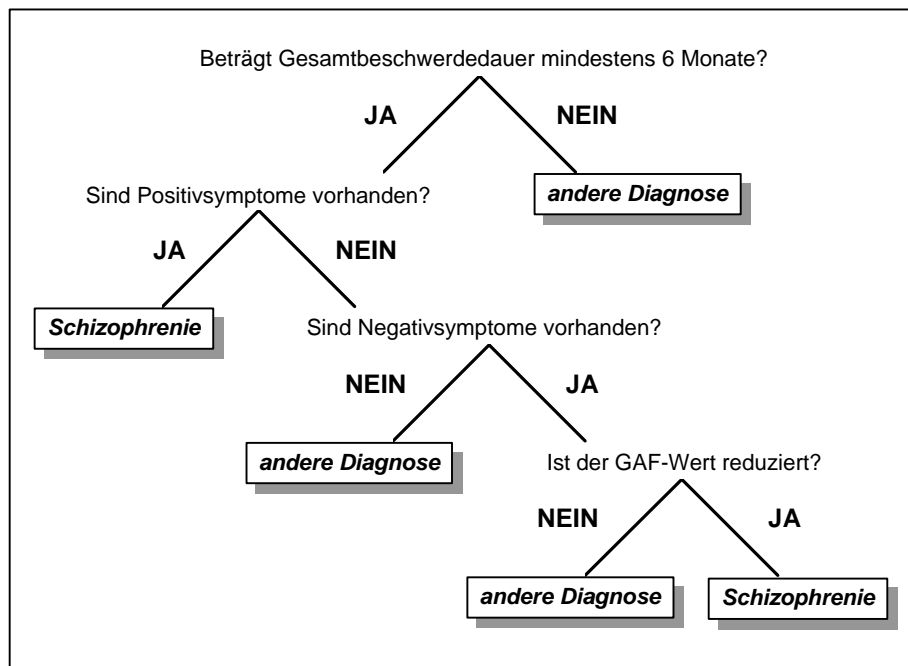


Abb. 19: Beispiel eines durch rekursive Partitionierung gewonnenen Dendrogramms zur Diagnose von Schizophrenie. Die in Kästchen gefaßten 'Endknoten' kennzeichnen die Klassen der abhängigen Variablen.

In Abbildung 19 ist das Ergebnis einer rekursiven Partitionierung am Beispiel einer grob vereinfachten Klassifizierung von psychiatrischen Patienten in solche mit der Diagnose einer Schizophrenie, $y = 1$, und solche mit einer anderen Diagnose, $y = 0$, in Form eines Dendrogramms wiedergegeben. Wenn die Attribute $X_1 = \{\text{Gesamtbeschwerdedauer} \geq 6 \text{ Monate}\}$, $X_2 = \{\text{Positivsymptome vorhanden}\}$, $X_3 = \{\text{Negativsymptome vorhanden}\}$ und $X_4 = \{\text{reduzierter GAF-Wert}\}$ bezeichnen, dann läßt sich das graphisch repräsentierte Partitionierungsergebnis für $y = 0$, 'andere Diagnose', formal folgendermaßen als Boolesche Funktion ausdrücken: $A_0 = (\bar{X}_1) \cup (X_1 \cap \bar{X}_2 \cap \bar{X}_3) \cup (X_1 \cap \bar{X}_2 \cap X_3 \cap \bar{X}_4)$. Eine andere Diagnose wird vergeben, wenn die Gesamtbeschwerdedauer weniger als sechs Monate beträgt, (\bar{X}_1) , oder wenn sie mindestens sechs Monate beträgt und weder Positiv- noch Negativsymptome vorhanden sind, $(X_1 \cap \bar{X}_2 \cap \bar{X}_3)$, oder wenn die Gesamtbeschwerdedauer mindestens sechs Monate be-

trägt, keine Positiv-, aber Negativsymptome vorhanden sind und keine Reduktion des globalen Funktionsniveaus, des GAF-Wertes, verzeichnet werden kann ($X_1 \cap {}^kX_2 \cap X_3 \cap {}^kX_4$).

Eine solche der graphischen Repräsentation entnommene Boolesche Funktion kann jedoch, wie auch in diesem Beispiel, redundante Information enthalten. Da der einzige Pfad zu $X_3 = \{\text{Negativsymptome vorhanden}\}$ über ${}^kX_2 = \{\text{keine Positivsymptome vorhanden}\}$ führt und der einzige Pfad zu ${}^kX_4 = \{\text{GAF-Wert nicht reduziert}\}$ über kX_2 und X_3 (s. Abb. 19), läßt sich A_0 vereinfacht auch als $A_0 = ({}^kX_1) \cup (X_1 \cap {}^kX_3) \cup (X_1 \cap {}^kX_4)$ ausdrücken: Eine andere Diagnose wird vergeben, wenn die Gesamtbeschwerdedauer weniger als sechs Monate beträgt, (kX_1), oder wenn sie mindestens sechs Monate beträgt und keine Negativsymptome vorhanden sind, ($X_1 \cap {}^kX_3$), oder wenn sie mindestens sechs Monate beträgt und der GAF-Wert nicht reduziert ist ($X_1 \cap {}^kX_4$).

Während in diesem Beispiel die in der Partition enthaltene redundante Information noch recht einfach zu erkennen ist, kann in komplexeren Dendrogrammen, in denen die zugrunde liegende Struktur der Partitionierung weniger offensichtlich ist, die Reduktion der Booleschen Funktion auf ihre einfachste und keine Redundanz enthaltene Form mit erheblichen Schwierigkeiten verbunden sein (Flegg, 1965).

In der Praxis stellt jedoch das Auftreten von komplementären Termen wie kX_3 oft ein größeres Problem als mögliche Redundanzen bei der Interpretation der Ergebnisse dar, da diese häufig den klinischen Apriori-Erwartungen entgegengesetzt sind (Marshall, 1995). Dieses Problem sei am Beispiel der Früherkennung von Schizophrenie erläutert. So forderten etwa Jackson und Mitarbeiter (1996) die Untersuchung aller möglichen Prodromalsymptome zur Verbesserung der Prädiktionsleistung von Früherkennungsinstrumenten. Dies basiert auf der Annahme, daß das Vorhandensein jedes dieser möglichen Prodromalsymptome positiv mit dem späteren Übergang in eine schizophrene Erkrankung assoziiert sein wird. Diesen Erwartungen entgegengesetzte Ergebnisse, etwa daß das Vorhandensein eines möglichen Prodromalsymptoms die Nicht-Entwicklung einer späteren Schizophrenie anzeigt bzw. die Abwesenheit eines Symptoms einen späteren Übergang in eine Schizophrenie indiziert, mögen deshalb auf den ersten Blick paradox erscheinen und zu einem Ausschluß des entsprechenden Symptoms verleiten. Ein solcher Ausschluß ist gerechtfertigt, wenn es sich bei dem entsprechenden Term um ein Artefakt handelt, das im Zuge der Partitionierungssequenzen entstanden ist. Doch sollte nicht von vornherein die Möglichkeit ausgeschlossen werden, daß eine aufgetretene Kombination von gleichzeitiger An- und Abwesenheit möglicher Prodromalsymptome auch klinisch sinnvoll sein kann. So wäre es bezogen auf das Basissymptom-Konzept (vgl. Kap. 2.1.7) beispielsweise vorstellbar, daß vorhandene eher eigenartige 'Stufe-2'-Denkstörungen wie Sprachstörungen oder Gedankenblockierungen sich als weniger prä-

diktiv für eine spätere Schizophrenie herausstellten, wenn sie ausschließlich zusammen mit 'Stufe-1'-Denkstörungen wie Konzentrationsstörungen oder Verlangsamung und Erschwerung von Denkprozessen auftreten. Durch eine entsprechende Kodierung von x_i , etwa auf der Basis von Korrelationsberechnungen, die nur positive Assoziationen zulässt, und eine a priori Restriktion von A_0 auf Boolesche Funktionen ohne komplementäre Terme lassen sich solche gemischten Ausdrücke jedoch auch leicht vermeiden, so daß der Vorannahme, daß das Vorliegen von Symptomen das Vorliegen einer Erkrankung und die Abwesenheit von Symptomen die Abwesenheit der Erkrankung anzeigt, entsprochen werden kann.

Neben diesen eher interpretativen Schwierigkeiten wurden vier weitere Argumente von Kritikern der rekursiven Partitionierung immer wieder vorgebracht (Doyle, 1973; Einhorn, 1972; Marshall, 1995):

1. die Notwendigkeit großer Stichproben zur Gewährleistung sinnvoller Subgruppengrößen,
2. die damit verbundene Ineffizienz im Umgang mit den Daten, da die Analyseeinheiten nicht die empirischen Daten, sondern die Untergruppen bilden,
3. das Fehlen von der Datenanalyse vorgeschalteten Modellannahmen und die daraus resultierende Anfälligkeit für die Generierung verfälschter Beziehungen und
4. das Vorhandenseins einer optimalen Gruppentrennung immer nur für bestimmte Untergruppen in der Hierarchie (lokales Optimum).

Dennoch ließ sich in verschiedenen Studien zeigen, daß die rekursive Partitionierung eine sinnvolle Analysemethode zur Erstellung von Klassifizierungsregeln für prognostische und diagnostische medizinische Zwecke darstellt (Armitage & Gehan, 1984; Breiman et al., 1984; Feinstein, 1977; Koss & Feinstein, 1971; Marshall, 1986, 1995; Marubini, Morabito & Valsecchi, 1983; Pascual, Rodriguez, Zayas, Moreno & Lage, 1983). Ein empirischer Vergleich der rekursiven Partitionierung mit konventionellen multivariaten inferenzstatistischen Prozeduren ermittelte eine ähnliche Güte der Verfahren (Cook & Goldman, 1984), aber es spricht nach Ansicht von Marshall (1986) für die Anwendung von Partitionierungstechniken, daß „partitioning is more intelligible to clinicians and has closer affinity to clinical logic.“ (ebd., p. 518).

Ein alternativer nicht-hierarchischer Partitionierungsansatz speziell für die Klassifizierung eines dichotomen Merkmals bzw. Ausgangs, der die Einschränkungen der 'klassischen' rekursiven Partitionierung vermeiden helfen soll, die 'Search Partition ANALysis - SPAN', wurde 1986 von Marshall vorgeschlagen. Anstelle der Analyse über die progressive Aufteilung in immer kleinere Gruppen wird hierbei eine direkte Suche in den gesamten Rohdaten mit dem Ziel vorgeschlagen, die beste unter einer bestimmten Klasse dichotomer Partitionierungen auszuwählen. Da dieser Ansatz zudem auf der Grundannahme gerichteter Zusammenhänge zwischen der abhängigen und den unabhängigen Variablen basiert, vermeidet er auch das Entstehen gemischter Boolescher Funktionen.

2.2.3.1 Search Partition Analysis

Wie die rekursive Partitionierung zielt auch die 'Search Partition ANalysis - SPAN' (Marshall, 1985, 1986, 1995) auf die Ermittlung Boolescher Funktionen ab, die entweder die Form einer Vereinigung (' \cup ') von Gruppen geschnittener Attribute X_i haben oder die einer Schnittmenge (' \cap ') vereinigter Gruppen von Attributen. Im Gegensatz zur rekursiven Partitionierung ist dies in SPAN mit der Restriktion verbunden, daß ausschließlich nicht-komplementäre oder ausschließlich komplementäre Attribute X_i gemeinsam in einer Funktion auftreten. Es werden also nur Partitionen zugelassen, die eine Vereinigung von q Gruppen geschnittener Mengen sind ($A_0 = I_1 \cup I_2 \cup \dots \cup I_q$, wobei I_s mit $s = 1, \dots, q$ die Schnittmenge einer Gruppe von p_s X_i s ist) oder eine Schnittmenge von q Gruppen vereinigter Mengen ($A'_0 = J_1 \cap J_2 \cap \dots \cap J_q$, wobei J_s die Vereinigung einer Gruppe von p_s X_i s ist). A_0 und A'_0 sind dabei alternative Partitionen, die sich zueinander wechselseitig verhalten, wobei A'_0 nicht die komplementäre Menge von A_0 ist. Die zu A_0 komplementäre Menge A_1 stellt zwar auch - wie A'_0 - die Schnittmenge von q vereinigten Mengen dar, diese werden jedoch aus den komplementären Attributen, $^kX_{i,s}$, gebildet. In SPAN wird q als „order“ (Ordnung) der Partitionen, p_1, p_2, \dots, p_q als ihre „subordinates“ (Marshall, 1986, p. 521) und A_0 und A'_0 als reguläre Partitionen bezeichnet.

Zur Vermeidung redundanter Informationen und damit zur Generierung möglichst einfacher Boolescher Funktionen wurde in SPAN darüber hinaus eine zweite Restriktion eingeführt, die im Falle von Funktionen der Form A_0 das Auftreten einer Schnittmenge I_s vermeidet, die bereits als Untermenge in einer anderen Schnittmenge I_r enthalten ist ($I_s \notin I_r$). Ist diese Voraussetzung erfüllt, wird keine Redundanz in A'_0 auftreten, da hiermit ebenfalls gesichert ist, daß keine Mengen vereinigt werden, die bereits als Untermenge in einer anderen vereinigten Menge vorhanden sind, $J_s \notin J_r$.

SPAN zielt hiermit darauf ab, die Stichproben-Gesamtmenge in zwei weitgehend getrennte Obermengen A_0 und A_1 zu unterteilen, die zwischen den alternativen Ausprägungen der abhängigen Variable diskriminieren und gleichzeitig durch die Beschreibung von A_0 und A_1 in einer regulären disjunktiven Form die Identifikation von Untermengen bzw. Subtypen und deren Überschneidungsbereich erlauben.

Über den in SPAN implementierten *Lock and Key*-Algorithmus (Marshall, 1995) kann eine direkte Suche nach der optimalen regulären Partition vorgenommen werden, wobei für eine gegebene Ordnung q und ihre Subordinaten p_q zunächst eine Aufstellung aller regulären Partitionen ('keys') für eine gegebene Anzahl von abhängigen Variablen x_p erfolgt, von denen diejenige ausgewählt wird, die am besten zwischen den beiden Ausprägungen von y diskriminiert, d.h. die größte 'Reinheit' ('lock') aufweist. Im Gegensatz zu der hierarchischen Methode der rekursiven Partitionierung, bei der die Daten immer weiter in immer kleinere

Subgruppen unterteilt werden, ermittelt SPAN jede Partition über die Gesamtdaten. Hierdurch wird die bei der hierarchischen Partitionierung auftretende Problematik der Notwendigkeit sehr großer Stichproben und der Ineffizienz der Analyseeinheiten vermieden. In Abhängigkeit von den Vorgaben in den Analyseschritten für q bei $p_1 = \dots = p_q = u$ ergeben sich $2(t \text{ über } q)$ mögliche Partitionen, wobei $t = (p \text{ über } u)$, da es t Möglichkeiten für jedes I_s (oder J_s) gibt, von denen q ausgewählt werden. Bei einer großen Anzahl von unabhängigen Variablen werden also sowohl p als auch q eher klein gewählt werden, um die Zahl der in den einzelnen Analyseschritten zu berücksichtigenden Partitionen nicht explosionsartig ansteigen zu lassen (Marshall, 1986).

Zur Bewertung der Güte gefundener Partitionen und damit der Auswahl der am besten zwischen dem Ausgang diskriminierenden kommen in SPAN zwei Meßkriterien zur Anwendung: zum einen die Messung der Partitionsreinheit, zum anderen die der Untergruppenreinheit („*partition purity*“ und „*subgroup purity*“, Marshall, 1995, p. 2651). Als Maß für die Reinheit der Gruppen wurde von Breiman und Mitarbeitern (1984) der Gini-Index G zur Schätzung der Verschiedenheit der Gruppen bzw. Partition über ihre Wahrscheinlichkeiten vorgeschlagen. Im Falle einer dichotomen abhängigen Variablen mit zwei Klassen ist G die Differenz zwischen dem Produkt der beiden beobachteten Wahrscheinlichkeiten der Klassen der abhängigen Variablen und den Produkten der einzelnen durch die Partitionen vorhergesagten Klassenwahrscheinlichkeiten. Die Reduktion der Verschiedenheit von y durch die Partitionierung in A_0 und A_1 kann für die Auswahl der 'besten' Partition genutzt werden:

$$G = \pi_0\pi_1 - p(A_0)p(0|A_0)p(1|A_0) - p(A_1)p(0|A_1)p(1|A_1),$$

wobei π_i die Wahrscheinlichkeit für $y = i$,
 $p(A_j)$ die Wahrscheinlichkeit für die Zugehörigkeit von x zu A_j
 $p(i|A_j)$ die bedingte Wahrscheinlichkeit für $y = i$ unter der Voraussetzung der Zugehörigkeit von x zu A_j für $i = 1, 0$ und $j = 1, 0$

Bei einer gegebenen Partition A_0 können die beobachteten Daten in eine 4-Felder-Tafel übertragen werden (s. Tab. 5). Unter der Voraussetzung, daß die Randsummen C und D in Tabelle 5 die wahren Anteile der beiden Ausprägungen von y in der Population, aus der die Berechnungstichprobe gewonnen wurde, widerspiegeln, vereinfacht sich der Gini-Index G durch die entsprechenden Ersetzungen der Wahrscheinlichkeiten durch die Zellenhäufigkeiten, etwa von π_0 durch C / N , von π_1 durch D / N , von $p(A_0)$ durch A / N und von $p(0|A_0)$ durch a / A , zu:

$$G = [(CD / N) - (ab / A) - (cd / B)] / N.$$

Damit hat G folgende Beziehung zu der 2×2 - χ^2 -Statistik: $G = \chi^2 CD / N$.

Ist die Annahme, daß die Randsummen C und D den real in der Population gegebenen Verhältnisse entsprechen, nicht haltbar, muß das Bayes Theorem mit Schätzung der Apriori-Wahrscheinlichkeiten π_i zur Berechnung von G herangezogen werden. Dabei ergibt sich un-

ter Verwendung der geschätzten verbundenen Wahrscheinlichkeiten anstelle der bedingten Wahrscheinlichkeiten (s. Tab. 5): $G = \pi_0\pi_1[1 - \{ab / (\pi_0aD + \pi_1bC)\} - \{cd / (\pi_0cD + \pi_1dC)\}]$.

Tab. 5: 2x2-Tabelle für N Probanden, klassifiziert nach einer binären abhängigen Variablen y mit den Ausprägungen 0 und 1 unter Verwendung einer 2-Klassen-Partition der unabhängigen Variablen (A₀ und A₁). Die in Klammern gesetzten Ausdrücke bezeichnen die geschätzten verbundenen Zellenwahrscheinlichkeiten für eine gegebene Apriori-Wahrscheinlichkeit $\pi_0 = p(y = 0)$ mit $\pi_1 = 1 - \pi_0$.

Partition der unabhängigen Variablen	Klassifizierung		
	y = 0	y = 1	Summe
A ₀	a (π_0a / C)	b (π_1b / D)	A = a + b
A ₁	c (π_0c / C)	d (π_1d / D)	B = c + d
Summe	C = a + c	D = b + d	N = a + b + c + d

Für die Praxis empfiehlt Marshall (1986) die Verwendung einer balancierten Partition, in der die Randsummen A und B annähernd gleich sind. Damit wird eine Gewichtung von G entsprechend der Disparität in A und B auf einfachem Wege ermöglicht, indem $G^* = GAB$ gesetzt wird, da $AB = A(N - B)$ maximal ist, wenn $A = N / 2$.

Alternativ wird für eine binäre abhängige Variable y die Messung der Entropie über das Verhältnis der Anzahl fälschlicher Zuordnungen zu A₀ zu der Gesamtzahl der Zuordnungen zu A₀, (b/A), vorgeschlagen (Marshall, 1995), wobei $i(A_0)$ den Meßwert für die Nicht-Reinheit der Partition A₀ darstellt, d.h. $i(A_0)$ wird kleiner mit zunehmender Partitionsreinheit:

$$i(A_0) = - (b/A) \log (b/A).$$

Ein weiteres Maß zur Berechnung der Nicht-Reinheit $i(A_0)$, das ebenfalls auf das Verhältnis b/A Bezug nimmt und sowohl für intervallskalierte als auch für binäre abhängige Daten verwandt werden kann, ist die Fehlervarianz:

$$i(A_0) = \sum (y_i - \bar{y}_{A_0})^2 / n_{A_0}, \text{ wobei } \bar{y}_{A_0} \text{ der Anteil von } y = 1 \text{ in } A_0 \text{ ist (b/A).}$$

Wird der Gini-Index G als Maß der Reinheit auf das Maß der Nicht-Reinheit $i(A_0)$ zurückgeführt, so lautet er: $G = i(A_0 \cup A_1) - p_{A_0}i(A_0) - p_{A_1}i(A_1)$, wobei p_{A_0} und $p_{A_1} = 1 - p_{A_0}$ der Anteil der Stichprobe in A₀ und A₁. G als Maß der Partitionsreinheit ist damit die Differenz zwischen der Nicht-Reinheit der verbundenen, zueinander komplementären Partitionen A₀ und A₁ und den an ihren jeweiligen Anteilen an der Stichprobe gewichteten einzelnen Nicht-Reinheiten, $i(A_0)$ und $i(A_1)$.

Zur Messung der Untergruppenreinheit $i(l_j)$, für die ebenfalls der Gini-Index herangezogen werden kann, werden $i(A_0)$ und $i(A_1)$ ersetzt durch den an der Gruppengröße gewichteten

Mittelwert jeder Untergruppenreinheit, $\bar{i}(A_0)$ und $\bar{i}(A_1)$, entsprechend der Formel:

$$\bar{i}(A_0) = [\sum_{j=1}^q i(l_j)n_j] / \sum_{j=1}^q n_j.$$

Wie CART enthält auch SPAN eine Straffunktion β („penalty parameter“, Marshall, 1995, p. 2652), mit deren Hilfe eine Relativierung der steigenden Reinheit an der meist damit verbundenen steigenden Komplexität der Partition c erfolgt: $G_\beta = G - c\beta$.

Für die Bestimmung der Komplexität c stehen verschiedene Vorgehensweisen zur Auswahl (van Mechelen & DeBoeck, 1990):

1. Die Komplexität bestimmt sich durch die Gesamtzahl der generierten Untergruppen:

$$c = q + q' - 1,$$

2. Die Komplexität c entspricht der Summe der Attribute in der Partition: $c =$
Anzahl X_i Diese Bestimmung ist weniger strikt als die vorherige, da es möglich ist, daß zwar insgesamt wenige Attribute heran gezogen werden, aber die einzelnen Ausdrücke für A_0 und A_1 dennoch recht kompliziert sind.
3. Die Komplexität entspricht der minimal möglichen Zahl von Subordinaten: $c = \min(q, q')$.
Unter dieser Vorgehensweisen wird eine möglichst einfache Boolesche Funktion zur Beschreibung nur einer der beiden Hauptgruppen gefordert.

Gemäß dem 'Lock and Key'-Algorithmus werden die generierten Partitionen in jedem Analyseschritt hinsichtlich G^* , d.h. der an der Disparität in A und B (s. Tab. 5) gewichteten Partitionsreinheit, bewertet und die Partition mit der höchsten Reinheit, dem höchsten G^* ausgewählt. Diese kann als eigenständige Variable abgespeichert und im nächsten Schritt verrechnet werden, wobei die hierfür ermittelte Straffunktion β den Voreinstellungswert für den nächsten Suchschritt dieser iterativen Suchstrategie bildet. Eine solche iterative Suchstrategie wird besonders bei einer großen Menge abhängiger Variablen empfohlen, denn bei $q > 2$ und $p_q > 3$ nimmt die Menge der zu durchsuchenden Kombinationsmöglichkeiten explosionsartig zu und führt damit zu immensen, unökonomischen Rechenzeiten (Marshall, 1995). Die iterative Suche wird abgebrochen, wenn sie konvergiert, d.h. wenn eine Erhöhung der Komplexität nicht mehr durch eine entsprechende Erhöhung der Partitionsreinheit aufgewogen wird. Dieses Abbruchkriterium und die damit verbundene Selektion von Items stellen einen weiteren Vorzug von SPAN gegenüber den hierarchischen Partitionierungsverfahren dar, die prinzipiell die Stratifizierung so lange fortsetzen, bis jedes Objekt einen eigenen Endpunkt bildet.

Da die Suche nach der optimalen Partition in SPAN auf die Menge der regulären Partitionen begrenzt ist, können artifizielle oder uninterpretierbare Attributkombinationen nicht entstehen, aber möglicherweise tatsächlich bestehende Interaktionen zwischen positiven und negativen Attributen nicht aufgedeckt werden. Generell werden jedoch meist unabhängige Variablen in

eine Studie aufgenommen werden, für die schon im Vorfeld Anhaltspunkte für einen Einfluß auf die abhängige Variable und dessen positive Richtung bestehen. Gibt es bereits Hinweise auf eine negative Beziehung zwischen der abhängigen und einer unabhängigen Variable, ist es jedoch auch möglich, diese im Vorfeld zuzulassen, indem etwa die Abwesenheit der unabhängigen Variable, $x_i = 0$, als positives Attribut definiert wird. Damit werden Kombinationen negativer und positiver Attribute in Form einer regulären Partition und die Prüfung jedweder Modellvorstellung ermöglicht.

Einen Vorzug von SPAN bietet die Definition der Untergruppen, die als intrinsisches Merkmal des regulären Booleschen Ausdrucks generiert werden. Hierbei kann das Ausmaß der Überlappung von Untergruppen einer Hauptmenge A_0 oder A_1 einen Eindruck des Ausmaßes der Verschiedenheit der in der Hauptgruppe enthaltenen Individuen geben: Eine starke Überlappung spricht hierbei für eine weitgehend homogene Hauptmenge, eine geringe Überlappung für eine heterogene, die aus mehreren unterscheidbaren Untergruppen gebildet wird.

Einen Nachteil von SPAN bildet jedoch die Tatsache, daß zur Zeit noch keine befriedigende Möglichkeit besteht, die statistische Signifikanz der generierten Partition zu ermitteln (Marshall, 1995)¹. Um zumindest die Güte der Gruppentrennung zu bestimmen, werden jedoch verschiedene Maße ermittelt, darunter χ^2 -Statistiken, positive und negative Stärke sowie Sensitivität und Spezifität.

Zwei Vergleichsstudien zu der Vorhersageleistung der generierten Entscheidungsregeln einer 'klassischen' rekursiven Partitionierung und der nicht-hierarchischen SPAN anhand von Daten zur Prädiktion eines geringen Geburtsgewichts bzw. zur Güte eines Glukosetoleranz-Screeningverfahrens ergaben etwa gleiche Vorhersagestärken für beide Verfahren, wobei SPAN jedoch eine leichte Überlegenheit vor allem in der positiven prädiktiven Stärke zeigte (Marshall, 1995).

2.2.4 Grob-Mengen-Theorie

Ein weiteres alternatives, nicht inferenzstatistisch basiertes Klassifizierungs- und Vorhersagemodell, das Variablen jeden Skalierungsniveaus simultan verarbeiten kann und speziell für die Datenanalyse, -reduktion und -modellierung sowie den Aufbau von Klassifikations-Experten-Systemen entwickelt wurde, ist das Grob-Mengen-Modell (engl.: „rough sets“) von Pawlak (1982, 1984, 1986; Pawlak, Wong & Ziarko, 1988). Es dient in erster Linie der De-

¹ Laut persönlicher Mitteilung des Autors lag eine solche auch zum Zeitpunkt des Entstehens der vorliegenden Arbeit noch nicht vor.

pendenzanalyse und Mustererkennung in Datentabellen, wobei es im Gegensatz zu Partitionierungsmodellen nicht auf die Generierung einzelner homogener Subgruppen abzielt, sondern auf die Entwicklung einfacher Entscheidungsregeln, anhand derer vorgegebene Gruppen erkannt werden können. Eine Reihe von Programmen wurde inzwischen für Berechnung grober Mengen entwickelt (ANLYST: Ziarko, 1987; LERS: Grzymala-Busse, 1992; PRIMEROSE: Tsumoto & Tanaka, 1995; R²: Hashemi, 1989; ROSETTA: Øhrn & Komorowski, 1997; Øhrn, Komorowski, Skowron & Synak, 1998; RoughDAS: Gruszecki, Slowinski & Stefanowski, 1990). Während einige dieser Programme das Grob-Mengen-Modell mit Elementen aus anderen Ansätzen verbinden, etwa aus Boolescher Logik (ROSETTA), ist das in Kanada entwickelte Programm DataLogic (REDUCT System Inc., 1991-1992, 1991-1994) eng an der ursprünglichen Grob-Mengen-Theorie von Pawlak (1982, 1984, 1986) orientiert.

Die Grob-Mengen sind ein mengentheoretisch fundiertes Modell, das der Approximations-Klassifikation der abhängigen Variable über die Generierung von Entscheidungsregeln für jede ihrer Ausprägungen im Sinne der Ermittlung spezifischer Datenmuster dient. Die Klassifikation betrifft eine Menge von Objekten (z.B. Basissymptome), die durch eine Menge von Attributen (z.B. Symptom vorhanden, Symptom nicht vorhanden) beschrieben sind. Jedem Paar von Objekt und Attribut entspricht ein Wert, der *Deskriptor*. Die Objekte, Attribute und Deskriptoren stellen die drei Basisbestandteile des Informationssystems, der Datentabelle, dar. Bei den Attributen sind Bedingungs- und Entscheidungsattribute zu unterscheiden. Erstere, auch Kausalattribute genannt, beschreiben die Struktur der Objekte, letztere, auch Konsektivattribute genannt, definieren die Klassifikation der Objekte nach den Werten der abhängigen Variable. Grundbegriffe der Grob-Mengen-Theorie sind die untere und obere Approximation, Schätzung, der Menge. Bei Anwendung einer bestimmten Teilmenge von Kausalattributen kann für jedes Objekt im Informationssystem bestimmt werden, ob es zur oberen oder unteren Schätzung der Menge entsprechend der gegebenen Klasse der unabhängigen Variablen gehört. Zählt das Objekt zu der unteren Approximation, ist es unzweifelhaft Teil dieser Klasse, gehört es zur oberen, kann es auch möglicherweise Teil der angrenzenden Klasse sein.

2.2.4.1 Grundlagen

Der Grob-Mengen-Ansatz basiert weitgehend auf der klassischen Mengentheorie, unterscheidet sich von dieser jedoch in einigen wesentlichen Punkten, die im folgenden beispielhaft dargestellt werden.

In der klassischen Mengentheorie ist X definiert als eine Untermenge von Objekten x_i der Gesamtmenge $U = \{x_1, x_2, \dots, x_i\}$. So sei beispielsweise die Gesamtmenge der Objekte U definiert durch eine Datenbank, die die Symptome verschiedener Patienten beinhaltet, und X sei die Untermenge der Patienten mit Schizophrenie (s. Abb. 20). In der Vorstellung der klassischen Mengentheorie hat ein Patient entweder die Diagnose einer Schizophrenie erhalten oder nicht. Dies wird ausschließlich durch die Diagnose selbst festgelegt. Gibt es keine Angaben zur Diagnose, ist in eine Zuordnung der Diagnose 'Schizophrenie' oder 'andere Diagnose' zu einem Patienten nicht möglich, d.h. ein Objekt ist nur bei Kenntnis der Ausprägung der abhängigen Variable y , dem Konsektivattribut v_y , klassifizierbar, eine Klassifikation anhand der Ausprägung der unabhängigen Variablen, der Kausalattribute v_v , ist nicht möglich.

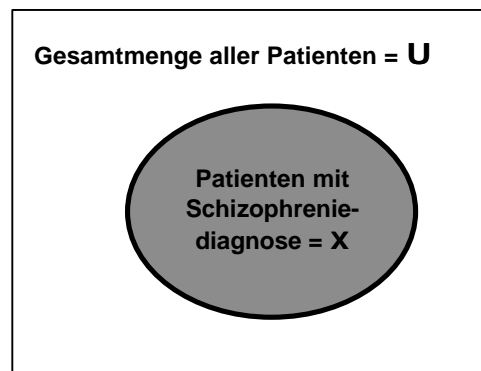


Abb. 20: Definition von Gesamtmenge und Untermenge nach der klassischen Mengentheorie

In den Modellvorstellungen der Grob-Mengen ist hingegen die Kenntnis der Klassifikationsgrundlagen, die Identifikation der durch die Konsektivattribute definierten Klassen anhand der Ausprägung der unabhängigen Variablen v_i , die die Menge V bilden, elementarer Bestandteil des Mengenmodells. So können hier bei Kenntnis der Symptome oder Kriterien, die mit einer schizophrenen Erkrankung assoziiert sind - etwa Wahn, Halluzinationen, formale Denkstörungen, Beschwerdedauer und globales Funktionsniveau - Patienten bzgl. ihrer Erkrankung klassifiziert werden, ohne die Diagnose zu kennen, indem die Beziehungen zwischen den vorliegenden Symptomen und der Diagnose zur Zuordnung herangezogen wird. Diese Beziehungen bilden das sogenannte *Klassifikationswissen* für die Erkrankung des Patienten, das in Form von Entscheidungsregeln ausgedrückt werden kann. Auf seiner Basis können die meisten Patienten den beiden Mengen 'Patienten mit Schizophrenie' und 'Patienten mit einer anderen Diagnose' zugeordnet werden. Diese Einbindung des Klassifikationswissens in die Mengentheorie ist die Hauptneuerung der Grob-Mengen-Theorie und die Grundlage zur Erstellung von Entscheidungs- bzw. Klassifikationsregeln.

Das Klassifikationswissen wird aus der Aufteilung der Gesamtmenge U in kleine unterscheidbare Kategorien, den sogenannte *elementaren Mengen* E gebildet. Eine elementare Menge beinhaltet Objekte der gleichen Ausprägungskombination einer oder mehrerer unabhängiger Variablen v_i (Kausalattribute v_v), die als Kriterium benutzt werden. Eine Unterscheidung zwischen den Objekten einer elementaren Menge anhand ihrer Attributausprägung ist nicht möglich, da sie im Hinblick auf ihre in einer elementaren Menge enthaltenen Kausalattribute absolut identisch sind. Damit bilden die elementaren Mengen von Objekten - basierend auf dem Klassifikationswissen - die kleinsten abzugrenzenden Objektmengen (s. Abb. 21). Hingegen beinhalten *Atome* \exists Objekte gleicher Ausprägungskombinationen aller Attribute. Sind alle Atome 1-Element-Mengen, ist das Informationssystem S , d.h. die Datentabelle, selektiv.

Das Klassifikationswissen wird formal durch die binäre Äquivalenzrelation $R(C)$ repräsentiert, die in einer Gesamtmenge von Objekten U definiert ist. Jede Äquivalenzrelation $R(C)$ entspricht einer Aufteilung von U in eine Anzahl elementarer Mengen, die durch eine Untermenge C aus der Menge aller unabhängigen Variablen V gebildet werden. Sind zwei Objekte x_i Elemente der gleichen Äquivalenzrelation $R(C)$, $(x_i, x_j) \in R(C)$, und besitzen sie die gleiche Informationsfunktion $f(x_i, v)$, die dem Objekt x_i jeweils eine Ausprägung pro unabhängiger Variable v zuordnet, $f(x_i, v) = f(x_j, v)$, und gilt dies für alle Objekte x_i der Gesamtmenge U , $x_j \in U$, und für alle unabhängigen Variablen v der Menge C , $v \in C$, dann sind die Objekte x_i und x_j im Näherungsraum A nicht unterscheidbar. Damit kann jedem Objekt x_i durch $R(C)$ eine einzigartige Menge V von Kausalattributen v_v zugeordnet werden, wobei C eine Untermenge der Menge aller unabhängigen Variablen V , maximal aber die gleiche Menge ist, $C \subseteq V$.

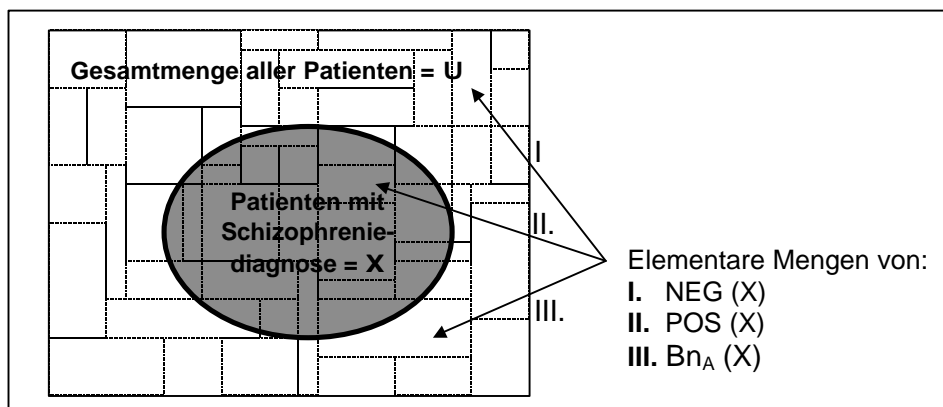


Abb. 21: Darstellung des Näherungsraums A

Die Aufteilung der Gesamtmenge U in einzelne elementare Mengen repräsentiert jeweils eine von mehreren möglichen Äquivalenzrelationen $R(C)$, wobei U und die mit ihr assoziierte

Äquivalenzrelation R Bestandteile des Näherungsraums A , $A = (U, R)$, sind (s. Abb.21). Der Näherungsraum A stellt somit eine angenäherte Beschreibung jeder Untermenge X von U unter der Annahme dar, daß auch leere Mengen elementar für jeden Näherungsraum A sind. Jede endliche Gruppe elementarer Mengen in A heißt *definierbare Menge*, $DEF(A)$, oder *Region* (s. Abb. 21 und 22). Die positive Region, $POS(X)$, ist die Gesamtheit aller elementaren Mengen, die vollständig in der Untermenge X liegen (s. Abb. 21 und 22). Objekte dieser Region können zweifelsfrei als der Untermenge X zugehörig eingestuft werden. Die negative Region, $NEG(X)$, ist die Gesamtheit aller elementaren Mengen, die vollständig außerhalb von X liegen. Objekte dieser Region können zweifelsfrei als nicht der Untermenge X zugehörig eingestuft werden. Elementare Mengen, die nur zum Teil mit der Untermenge X überlappen, bilden die Grenzregion, Bn_A (s. Abb. 21 und 22). Da sich Objekte einer elementaren Menge nicht voneinander unterscheiden lassen, ist das Klassifikationswissen einer Grenzmenge nicht ausreichend, um zu entscheiden, ob ein Objekt dieser Menge zur Untermenge X gehört oder nicht. Die Grenzregion $Bn_A(X)$ einer Untermenge X hinsichtlich $R(C)$ ist somit die Gesamtheit aller elementaren Mengen von A , die teilweise, aber nicht vollständig in X eingeschlossen sind. Nur in dem Fall, daß die Grenzregion $Bn_A(X)$ keine elementaren Mengen enthält, also eine leere Menge ist, $Bn_A(X) = \emptyset$, wäre das Klassifikationswissen perfekt und alle Objekte könnten zweifelsfrei als der Untermenge X zugehörig bzw. nicht zugehörig eingestuft werden.

Da die Grenzregion in den meisten Fällen jedoch keine leere Menge ist, sondern elementare Mengen enthalten wird, $Bn_A(X) \neq \emptyset$, wird durch die Spezifizierung der positiven, negativen und Grenzregion durch die Äquivalenzrelation $R(V)$ die Untermenge X der Gesamtmenge U nur annäherungsweise beschrieben, so daß X als eine sogenannte *grobe Menge* (engl.: „rough set“) definiert ist.

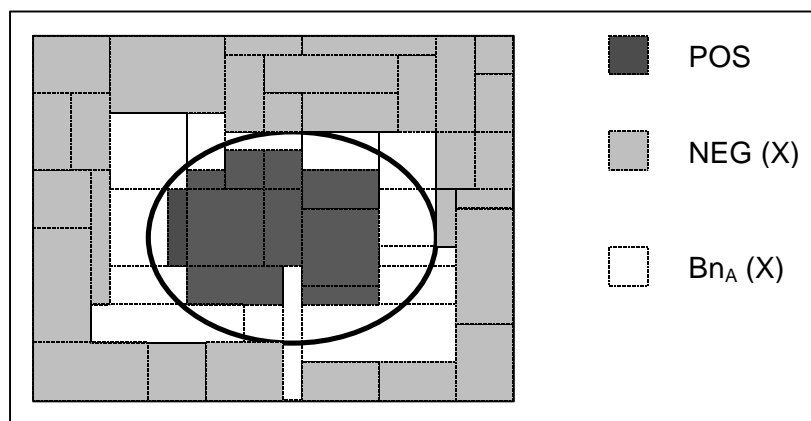


Abb. 22: Darstellung der verschiedenen Regionen im Näherungsraum A

Die positive Region POS (X) stellt für sich genommen die untere Schätzung der Untermenge X im Näherungsraum A dar, $POS(X) = \underline{A}X$. Sie enthält alle Objekte, die zweifelsfrei X zugehörig sind. Die Verbindung von positiver und Grenzregion stellt die obere Schätzung der Untermenge X im Näherungsraum A dar: $POS(X) \cup Bn_A(X) = \overline{A}X$. Sie beinhaltet alle Objekte, die möglicherweise X zugehörig sind. Somit lassen sich auch die Grenzregion und die negative Region NEG (X) durch die untere und obere Schätzung der Untermenge X im Näherungsraum A der Gesamtmenge U beschreiben: $Bn_A(X) = \overline{A}X - \underline{A}X$ und $NEG(X) = U - \overline{A}X$. Die Grenzregion ist gleich der Differenz von oberer und unterer Schätzung der Untermenge X, die negative Region hingegen die Differenz der Gesamtmenge U und der oberen Schätzung von X, die die Vereinigung der positiven und der Grenzregion darstellt.

Die obere und untere Schätzung, \overline{A} und \underline{A} , sind die Einschluß- bzw. Ausschlußoperationen im topologischen Raum T_A , der durch den Näherungsraum A definiert wird, $T_A = (U, DEF(A))$, und durch die folgenden mengentheoretischen Eigenschaften charakterisiert:

1. $\underline{A}X \subset X \subset \overline{A}X$ Die untere Schätzung der Menge X ist eine Untermenge von X, die eine Untermenge der oberen Schätzung von X darstellt.
2. $\underline{A}U = \overline{A}U = U$ Die untere Schätzung der Gesamtmenge entspricht ihrer oberen Schätzung bzw. der Gesamtmenge selbst.
3. $\underline{A}\emptyset = \overline{A}\emptyset = \emptyset$ Die untere Schätzung einer leeren Menge entspricht ihrer oberen Schätzung bzw. ist eine leere Menge.
4. $\overline{A}(X \cup Y) = \overline{A}X \cup \overline{A}Y$ Die obere Schätzung der vereinigten Untermengen X und Y ist gleich der Vereinigung von der oberen Schätzung von X mit der oberen Schätzung von Y.
5. $\underline{A}(X \cup Y) \supset \underline{A}X \cup \underline{A}Y$ Die untere Schätzung der vereinigten Untermengen X und Y ist eine Obergruppe der Vereinigung der unteren Schätzung von X und der unteren Schätzung von Y.
6. $\overline{A}(X \cap Y) = \underline{A}X \cap \underline{A}Y$ Die obere Schätzung der Schnittmenge von X und Y ist gleich der Schnittmenge der jeweils unteren Schätzungen von X und Y.
7. $\underline{A}(X \cap Y) \subset \underline{A}X \cap \underline{A}Y$ Die untere Schätzung der Schnittmenge von X und Y ist eine Untermenge der Schnittmenge der jeweils unteren Schätzung beider Mengen.
8. $\overline{A}({}^kX) = {}^k\underline{A}X$ Die obere Schätzung der komplementären Menge von X, kX , ist gleich dem Komplement der unteren Schätzung von X, ${}^k\underline{A}X$.
9. $\underline{A}({}^kX) = {}^k\overline{A}X$ Die untere Schätzung der komplementären Menge von X, kX , ist gleich dem Komplement der oberen Schätzung von X, ${}^k\overline{A}X$.

Damit ist eine Menge X in A nur dann vollständig definierbar wenn $\underline{A}X = \overline{A}X$, ist dies nicht der Fall, gehört X zu einer der folgenden vier Gruppen mit unterschiedlichen Graden an 'Nicht-Definierbarkeit':

1. $\underline{A}X \neq \emptyset$ und $\overline{A}X \neq U$ Wenn die untere Schätzung von X keine leere Menge ist und die obere Schätzung nicht gleich der Gesamtmenge U ist, dann ist X *grob definierbar* in A und kann geschätzt werden.
2. $\underline{A}X \neq \emptyset$ und $\overline{A}X = U$ Wenn die untere Schätzung von X keine leere Menge ist und die obere Schätzung gleich der Gesamtmenge U ist, dann ist X *extern undefinierbar* in A , und kein Attribut kann eliminiert werden.
3. $\underline{A}X = \emptyset$ und $\overline{A}X \neq U$ Wenn die untere Schätzung von X eine leere Menge und die obere Schätzung nicht gleich der Gesamtmenge U ist, dann ist X *intern undefinierbar* in A ; kein Element kann eindeutig X zugeordnet werden.
4. $\underline{A}X = \emptyset$ und $\overline{A}X = U$ Wenn die untere Schätzung von X eine leere Menge ist und die obere Schätzung gleich der Gesamtmenge U , dann ist X *vollständig undefinierbar* in A , und nicht einmal Schätzungen von X können definiert werden.

Ist die Untermenge X grob definierbar in A und kann damit geschätzt werden, so bestimmt sich die Genauigkeit der Untermengenschätzung, $\mu_A(X)$, durch das Verhältnis der in der oberen und unteren Schätzung eingeschlossenen Objekte:

$$\mu_A(X) = (\text{Anzahl der Objekte in } \underline{A}X) / (\text{Anzahl der Objekte in } \overline{A}X)$$

Hierbei kann $\mu_A(X)$ Werte von 0 bis 1 annehmen. Bei einem Wert von 1 ist die Schätzung eindeutig, da hierbei keine Grenzregion existiert und X vollständig definierbar in A ist, $\underline{A}X = \overline{A}X$. Da dies nur selten der Fall sein wird, ist die zentrale Aufgabe in des Grob-Mengen-Modells das Erstellen einfacher, nicht-redundanter Mengenbeschreibungen bei gleichzeitigem Erhalt der Genauigkeit der Mengenschätzung. Um dies zu erreichen, muß die Gesamtmenge U so in elementare Mengen eingeteilt werden, daß keine Trennung zwischen den elementaren Mengen aufgehoben werden kann, ohne die Genauigkeit der Schätzung von X zu verringern. Ein Software-Programm, das dies leistet, ist DataLogic (REDUCT System Inc., 1991-1994).

2.2.4.2 DataLogic

Das Informationssystem S , das die gesamten Informationen über alle Objekte U , die abhängige und die unabhängigen Variablen enthält, $S = (U, V, V_v, f)$, bildet die Grundlage für die

Ermittlung der Klassifizierungsregeln durch die Identifizierung wiederkehrender und spezifischer Datenmuster mit einer möglichst kleinen Anzahl von unabhängigen Variablen in Data-Logic, das sich eng an der ursprünglichen Theorie von Pawlak (1982, 1984; Pawlak, Wong & Ziarko, 1988) anlehnt. In Form einer Datentabelle enthalten hierbei die Zeilen die Informationen über reale Objekte x , die Spalten Informationen über die Kausalattribute v_v (s. Tab. 6). Jede unabhängige Variable besitzt eine Anzahl möglicher Ausprägungen, den Bereich V_v : $V_v = \{\text{Ausprägung1, Ausprägung2, Ausprägung3, \dots, Ausprägung n}\} = \{v_1, v_2, v_3, \dots, v_v\}$.

Die Informationen über ein Objekt x_i ist in der Informationsfunktion $f(x_i, v)$ enthalten, die dem Objekt x_i jeweils eine Ausprägung pro unabhängiger Variable v zuordnet. Jedes aus der Gesamtmenge U stammende Objektpaar, (x_i, x_j) mit $x_i, x_j \in U$, das durch Kausalattribute v_v aus einer Menge von unabhängigen Variablen charakterisiert ist, $v \in C$, ist eine Beschreibung in S . Auf dieser Basis ist es möglich, mit jeder Untermenge C der Menge der unabhängigen Variablen V ($C \subseteq V$) einen Näherungsraum A in der Datentabelle durch die Definition einer Äquivalenzrelation $R(C)$ über die Gesamtmenge U wie folgt zu assoziieren:

$$(x_i, x_j) \in R(C), \quad \text{wenn } f(x_i, v) = f(x_j, v) \quad \text{für alle } x_i, x_j \in U \text{ und } v \in C.$$

Hierbei werden durch $R(C)$ die Objekte entsprechend der Anzahl der Ausprägungen der unabhängigen Variablen in C elementaren Mengen zugeordnet. Tabelle 6 verdeutlicht dies beispielhaft anhand der gleichen Informationen, die schon zur Illustration der rekursiven Partitionierung (vgl. Kap. 2.2.3, Abb. 19) herangezogen wurden (Symptom, Gesamtbeschwerdedauer und GAF-Wert als unabhängige und Diagnose als abhängige Variable).

Tab. 6: Beispiel eines Informationssystems S

U	V			
		(C)		(Y)
Patient	Symptom	Gesamtbeschwerdedauer	GAF-Wert	Diagnose
P1	Negativsympt.	6 Monate	reduziert	Schizophrenie
P2	Positivsympt.	> 6 Monate	reduziert	Schizophrenie
P3	Negativsympt.	< 6 Monate	normal	andere
P4	Positivsympt.	> 6 Monate	reduziert	Schizophrenie
P5	Positivsympt.	6 Monate	reduziert	Schizophrenie
P6	Positivsympt.	< 6 Monate	reduziert	andere
P7	Negativsympt.	< 6 Monate	erhöht	andere
P8	Negativsympt.	> 6 Monate	normal	andere
P9	Negativsympt.	< 6 Monate	normal	andere

Hierin ist:

$U = \{P1, P2, \dots, P9\}$

$V = \{\text{Symptom, Gesamtbeschwerdedauer, GAF-Wert, Diagnose}\}$

$V_{\text{Symptome}} = \{\text{Positivsymptome, Negativsymptome}\}$

$V_{\text{Gesamtbeschwerdedauer}} = \{\text{< 6 Monate, 6 Monate, > 6 Monate}\}$

$V_{\text{GAF-Wert}} = \{\text{reduziert, normal, erhöht}\}$

$V_{\text{Diagnose}} = \{\text{Schizophrenie, andere Diagnose}\}$.

$f(x, v)$ muß jeweils dem grau unterlegten Bereich der Tabelle entnommen werden.

Das in Tabelle 6 dargestellte Beispiel von neun Patienten P enthält sieben Atome \exists von S, d.h. sieben verschiedene Untermengen von Objekten identischer Attributausprägungen bzw. identischer elementarer Mengen: $\exists 1 = \{P1\}$, $\exists 2 = \{P2, P4\}$, $\exists 3 = \{P3, P9\}$,

$$\exists 4 = \{P5\}, \quad \exists 5 = \{P6\}, \quad \exists 6 = \{P7\}, \quad \exists 7 = \{P8\}.$$

Wird nun beispielsweise als Untermenge C der Menge der unabhängigen Variablen V die Gesamtbeschwerdedauer, $C = \{\text{Gesamtbeschwerdedauer}\}$, herausgegriffen, dann besitzt die Äquivalenzrelation $R(C)$ entsprechend den hierin enthaltenen Kausalattributen *< 6 Monate*, *6 Monate* und *> 6 Monate* drei elementare Mengen E, aufgeteilt nach:

$$E1 = \{P3, P6, P7, P9\} \quad < 6 \text{ Monate}$$

$$E2 = \{P1, P5\} \quad 6 \text{ Monate}$$

$$E3 = \{P2, P4, P8\} \quad > 6 \text{ Monate}$$

Für diesen Näherungsraum A kann nun die Schätzung der aus der Menge der Attribute gewählten Entscheidungsvariable Y, hier $Y = \{\text{Diagnose}\}$ berechnet werden.

Ist die Untermenge X das Attribut Diagnose = {Schizophrenie}, dann enthält sie nach Tabelle 6 vier Objekte, $X = \{P1, P2, P4, P5\}$.

Da E2 die einzige elementare Menge von $R(C)$ ist, deren Objekte alle in X enthalten sind, gilt: $\text{POS}(X) = \underline{A}X = E2 = \{P1, P5\}$.

Die Schnittmengen aller elementaren Mengen mit X sind:

$$E1 \cap X = \emptyset \quad \text{da kein Objekt in E1 auch ein Objekt in der Menge X ist,}$$

$$E2 \cap X \neq \emptyset \quad \text{wobei E2 eine Untermenge und damit eine positive Region von X ist,} \\ E2 \subset X, \text{ und}$$

$$E3 \cap X \neq \emptyset \quad \text{wobei die Schnittmenge nur zwei der drei Objekte von E3 enthält und} \\ \text{somit die Grenzregion darstellt, } (E3 \cap X) = \{P2, P4\}.$$

Die obere Schätzung ergibt sich aus dem Zusammenschluß der Mengen, die Objekte aus X enthalten, bzw. der positiven und der Grenzregion, also der elementaren Mengen E2 und E3:

$$\overline{A}X = E2 \cup E3 = \{P1, P2, P4, P5, P8\}$$

Da $\underline{A}X \neq \overline{A}X$, kann X nicht vollständig definiert, sondern nur grob mit einer Genauigkeit von $\underline{A}X / \overline{A}X = 2/5 = 0,4 = \mu_A(X)$ geschätzt werden.

Aus den Bestimmungen der oberen und unteren Schätzung von X lassen sich die in der Grenzregion und der negative Region befindlichen Objekte bestimmen:

$$Bn_A(X) = \overline{A}X - \underline{A}X = \{P2, P4, P8\} \quad \text{und}$$

$$NEG(X) = U - \overline{A}X = \{P3, P6, P7, P9\}$$

Bei der Generierung von Vorhersagemodellen auf der Basis wiederkehrender Datenmuster wird nun aber in den seltensten Fällen die Auswahl von nur einer unabhängigen Variablen für die Erstellung von möglichst präzisen Entscheidungsregeln ausreichend sein, so daß die

Identifizierung von Abhängigkeiten zwischen verschiedenen Variablengruppen notwendig wird.

Ist C eine Menge von Kausalvariablen und Y die Entscheidungsvariable und werden mehrere Variablen in C zur Aufteilung der Gesamtmenge in elementare Mengen genutzt, definiert dies die Äquivalenzrelation R(Y) mit $Y' =$ Gruppe der elementaren Mengen, die der Äquivalenzrelation R(Y) zugehörig sind. Hierbei wird also nicht nur eine, sondern es werden mehrere Äquivalenzrelationen ermittelt und mit ihnen zusammen die positive Region jeder Kategorie von Y, mit POS (C, Y) als Gesamtheit der positiven Regionen der Aufteilung Y' :

$$\text{POS}(C, Y) = \bigcup \{R(C) Y : Y \in Y'\}$$

Wie im Falle nur einer herangezogenen unabhängigen Variable ist auch bei Einbezug mehrerer Variablen die positive Region von Y' erneut als ein eindeutig abgegrenzter Teil der Gesamtmenge U definiert, in dem jedes Objekt in POS (C, Y) durch die Kausalattribute der Menge C eindeutig einer der elementaren Mengen von Y' zugeordnet werden kann. Dabei besteht eine Abhängigkeit der Menge der Entscheidungsattribute Y von der Menge der Kausalattribute C im Ausmaß K, dem Verhältnis der Anzahl der Objekte in allen positiven Regionen der Entscheidungsattribute zu der Anzahl aller Objekte, mit $0 \leq K \leq 1$:

$$K = (\text{Anzahl der Objekte in POS}(C, Y)) / (\text{Anzahl der Objekte in der Gesamtmenge } U).$$

Ist $K = 1$ besteht eine vollständige Abhängigkeit: Alle Kategorien von Y' können durch die Attribute in C diskriminiert werden. Ist $K = 0$ besteht eine vollständige Unabhängigkeit: Durch die Attribute in C kann kein Objekt in irgendeine Kategorie in Y' klassifiziert werden. Nimmt K Werte zwischen 0 und 1 an, $0 < K < 1$, besteht eine partielle Abhängigkeit: Nur einige Objekte können durch die Attribute in C in Kategorien von Y' klassifiziert werden.

Angewandt auf das Beispiel in Tabelle 6 bedeutet dies: Bei $C = \{\text{Symptom, Gesamtbeschwerdedauer, GAF-Wert}\}$ und $Y = \{\text{Diagnose}\}$ ist für das Konzept *Schizophrenie* $K = 4/4 = 1$, d.h. die Abhängigkeit ist vollständig; jeder Patient, kann eindeutig als Person mit *Schizophrenie* oder *anderer Diagnose* klassifiziert werden.

Entscheidungsregeln sollen jedoch nicht nur möglichst präzise Objekte klassifizieren, sondern darüber hinaus auch möglichst ökonomisch sein, d.h. möglichst wenig Variablen enthalten. Da es wahrscheinlich ist, daß einige Attribute der Datentabelle redundant und/oder irrelevant sind und keinen Effekt auf die Diskriminationsleistung des Systems haben, sollte es im Zuge der Modellgenerierung möglich sein, diese Variablen zu identifizieren und ohne essentiellen Informationsverlust zu eliminieren. Dabei gilt es bei Erhalt der Attributabhängigkeit K eine Untermenge von Variablen von minimaler Größe zu ermitteln, das sogenannte *Attributredukt* M, an das die folgenden zwei Bedingungen gestellt werden:

1. Der Abhängigkeitsgrad K zwischen den Mengen C und Y muß erhalten bleiben.
2. Eine weitere Eliminierung von Attributen aus M ist ohne einen Abhängigkeitsverlust, einer Senkung von K , nicht möglich, d.h. M enthält keine weiteren Variablen mit redundanter Information.

Besitzt ein Informationssystem mehr als ein Redukt, kann jedes dieser Redukte anstelle der Menge der Bedingungsvariablen C verwandt werden. Dies sei erneut an den Schizophrenie-Daten aus Tabelle 6 erläutert: Eine Betrachtung der Objekte mit dem Entscheidungsattribut *Schizophrenie* zeigt, daß sie sich durch das Redukt $M1 = \{\text{Gesamtbeschwerdedauer, GAF-Wert}\}$ und durch das Redukt $M2 = \{\text{Symptome, Gesamtbeschwerdedauer}\}$ ohne Abhängigkeitsverlust beschreiben lassen. Aus diesen Redukten resultieren somit zwei Entscheidungsregeln für $M1$:

1. Wenn *Gesamtbeschwerdedauer = 6 Monate & GAF-Wert = reduziert*, dann *Diagnose = Schizophrenie*
2. Wenn *Gesamtbeschwerdedauer = > 6 Monate & GAF-Wert = reduziert*, dann *Diagnose = Schizophrenie*

und folgende drei Entscheidungsregeln für $M2$:

1. Wenn *Gesamtbeschwerdedauer = 6 Monate & Symptome = Negativsymptome*, dann *Diagnose = Schizophrenie*
2. Wenn *Gesamtbeschwerdedauer = 6 Monate & Symptome = Positivsymptome*, dann *Diagnose = Schizophrenie*
3. Wenn *Gesamtbeschwerdedauer = > 6 Monate & Symptome = Positivsymptome*, dann *Diagnose = Schizophrenie*

Entsprechend können auch Regeln für die negative Region von X (hier: $y_1 = \text{Schizophrenie}$) erstellt werden. Sind diese erfüllt, dann ist die Entscheidungskategorie 'nicht X ' (hier: $y_2 = \text{andere Diagnose}$). Solche aus POS (X) und NEG (X) abgeleiteten Regeln sind deterministisch, da nur jeweils ein Ergebnis mit ihnen assoziiert ist. Existiert jedoch - wie dies meist der Fall sein wird - eine Grenzregion, gibt es auch nicht-deterministische Regeln, mit mehr als einem Ergebnis. Am Beispiel von Tabelle 6 könnte eine solche nicht-deterministische Regel der Grenzregion lauten: Wenn *GAF-Wert = reduziert*, dann *Diagnose = Schizophrenie oder andere*.

Für die meisten empirischen Daten werden unpräzise Muster, d.h. das Vorkommen einer Grenzregion, deren Zellen teilweise eine große Anzahl von Objekten der Menge X enthalten, unvermeidbar sein. Diese Tatsache führt bei einer ausschließlichen Nutzung der positiven Region POS (X) bei der Beschreibung von X zu einem Verlust der Informationen über Objekte von X , die in der Grenzregion enthalten sind. Es ist deshalb wichtig, auch die in der

Grenzregion $B_{n_A}(X)$ enthaltenen Informationen zu analysieren. Obwohl die Mengenzugehörigkeit eines in $B_{n_A}(X)$ befindlichen Objekts x_i nicht bestimmt werden kann, ist es doch möglich den Anteil der Objekte x_i einer elementaren Menge von $B_{n_A}(X)$ zu ermitteln und somit die Wahrscheinlichkeit p zu schätzen, mit der ein in der Grenzregion befindliches Objekt x_i zu X gehört, $p(x_i \in X)$. Die Einführung von Wahrscheinlichkeiten erlaubt somit die Nutzung auch unpräziser Muster zur Beschreibung von X und gibt ein Maß für die Genauigkeit dieser Muster, wenn sie zur Klassifizierung von x_i herangezogen werden. Das Ausmaß dieser Genauigkeit kann über die Spezifizierung einer Schwellenwert-Wahrscheinlichkeit im Vorfeld der Analyse bestimmt werden. Elementare Mengen der Grenzregion, deren Wahrscheinlichkeit über diesem Schwellenwert liegt, werden dann POS (X) zugerechnet und zur Beschreibung von X herangezogen, die verbleibenden elementaren Mengen werden NEG (X) hinzugefügt. Abbildung 23 verdeutlicht dieses Vorgehen graphisch für eine vorgegebene Schwellenwert-Wahrscheinlichkeit von $p = 0,5$.

Durch die Wahl von $p < 1$ wird POS (X) so verändert, daß einige Information redundant wird und zusätzliche Variablen bei der Beschreibung von M entfallen. Es resultieren stärkere und generellere Entscheidungsregeln, die jedoch X nicht mehr mit hundertprozentiger Genauigkeit beschreiben. Dennoch wird die Verwendung solcher unpräzisen Muster als ein effektiver Weg betrachtet, um auch in stark von Nebeneffekten überlagerten, unpräzisen und inkonsistenten Daten starke Muster aufdecken zu können (Hashemi, Jelovsek & Razzaghi, 1993; Pawlak, Wong & Ziarko, 1988). Um dies zu ermöglichen, kann in DataLogic sowohl die 'rule roughness' (Regel-Grobheit) als auch der 'rule precision threshold' (Schwellenwert der Regel-Genauigkeit) für die Generierung von Entscheidungsregeln a priori festgelegt werden, die - ausgedrückt als Wahrscheinlichkeiten - Werte zwischen 1 und 0 annehmen können.

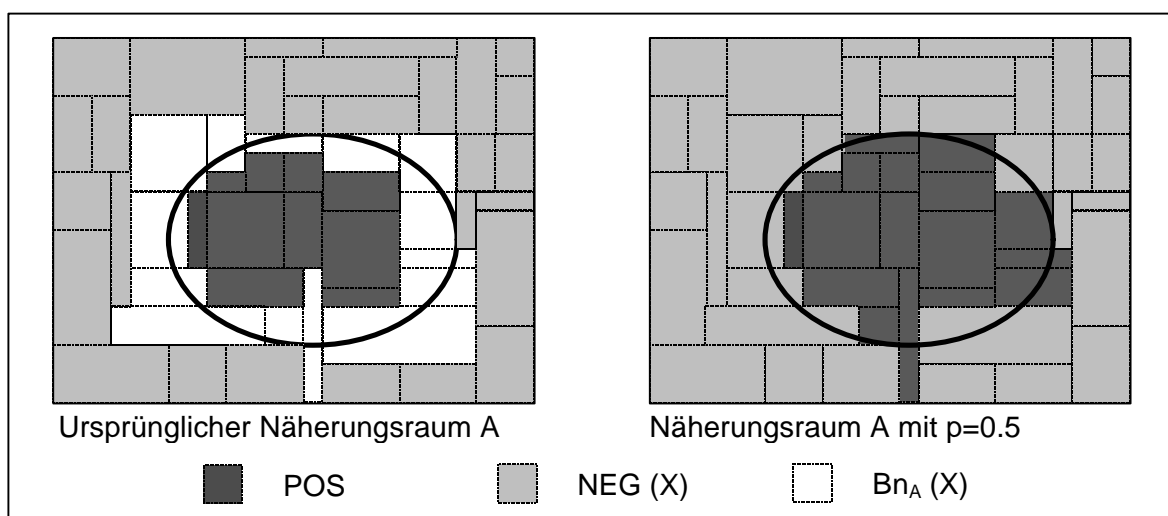


Abb. 23: Nutzung unpräziser Muster zur Beschreibung von X . Einführung eines Schwellenwertes für $p(x_i \in X)$ für $x_i \in B_{n_A}(X)$ und resultierende Neuordnung der elementaren Mengen.

Bei einer Vorgabe von 1 bei der rule roughness werden sehr generelle, 'grobe' Regeln generiert, mit sinkendem Voreinstellungswert tendieren die Regeln zu immer größerer Detailgenauigkeit, wobei ihre Anzahl steigt, da die einzelnen Regeln immer weniger Objekte beschreiben. Der rule precision threshold entspricht der geringsten Wahrscheinlichkeit, daß ein durch eine zutreffende Entscheidungsregeln klassifiziertes Objekt auch tatsächlich dem Bedingungsattribut y_v der Entscheidungsregel entspricht, d.h. bei einer Voreinstellung von 0.5 entsprechen mindestens 50% der den Bedingungen einer Entscheidungsregel entsprechenden Fälle tatsächlich dem vorhergesagten Ausgang.

Damit bietet das Grob-Mengen-Modell die Möglichkeit, generelle Muster in den individuellen Datenkonstellationen aufzudecken und dies unabhängig von inferenzstatistischen oder verteilungsbezogenen Vorannahmen. Da individuelle Unterschiede zudem nicht 'herausgemittelt' werden, können prinzipiell auch seltene, aber hoch prädiktive Zusammenhänge erkannt werden, die sich in inferenzstatistischen Verfahren möglicherweise in Form von 'Ausreißern' niederschlagen und aus der Analyse ausgeschlossen würden.

2.2.5 Diagnostische Effizienzindizes

Neben inferenzstatistischen und alternativen mathematischen Verfahren zur Generierung von Vorhersagemodellen bei gleichzeitiger Variablenselektion kommt in der Medizin häufig eine eher deskriptive Strategie der Variablenselektion zur Anwendung, die Auswahl einzelner Symptome nach ihrer diagnostischen Güte. Hierbei werden in Anlehnung an die quantitative Abschätzung der diagnostischen Güte medizinischer Testverfahren für einzelne qualitative, insbesondere binär-nominalskalierte Symptome sogenannte *diagnostische Effizienzindizes* (engl.: „indices of diagnostic efficiency“; Jackson, McGorry & Dudgeon, 1995, p. 244) herangezogen: Sensitivität, Spezifität, positive prädiktive Stärke (PPP) und negative prädiktive Stärke (NPP), die sich alternativ als Wahrscheinlichkeiten p mit Werten zwischen 0 und 1 (s. Tab. 7) oder in Prozentwerten mit Werten zwischen 0% und 100% ausdrücken lassen. Obwohl die diagnostischen Effizienzindizes ursprünglich für den Vergleich von Testverfahren entwickelt wurden (Begg & Greenes, 1983; Boyko, 1994; Leisenring & Pepe, 1998; McNeil, Keeler & Adelstein, 1975; Sackett, 1991), wurde ihre Anwendung später auch auf die Beschreibung der diagnostischen Güte einzelner Symptome ausgedehnt (Flaum & Andreasen, 1991; Jackson, McGorry & Dudgeon, 1995; Klosterkötter, Hellmich, & Schultze-Lutter, 2000; Olin & Mednick, 1996).

Die formalen Definitionen der einzelnen Effizienzindizes, die sich aus den unterschiedlichen Relationen von Zellbesetzungen in Kontingenztafeln ergeben, etwa einer 4-Felder-Tafel mit y

= {Erkrankung aufgetreten; Erkrankung nicht aufgetreten} und $x_i = \{\text{Symptom vorhanden; Symptom nicht vorhanden}\}$, sind in Tabelle 7 wiedergegeben.

Tab. 7: Definitionen der diagnostischen Effizienzindizes zur quantitativen Bestimmung der diagnostischen Güte binärer Merkmale als Wahrscheinlichkeiten p .

Indize	Definition
Sensitivität	$= p(\text{Symptom} \mid \text{Erkrankung})$
Spezifität	$= p(\neg \text{Symptom} \mid \neg \text{Erkrankung})$
positive prädiktive Stärke (PPP)	$= p(\text{Erkrankung} \mid \text{Symptom})$
negative prädiktive Stärke (NPP)	$= p(\neg \text{Erkrankung} \mid \neg \text{Symptom})$

Symptom: Symptom vorhanden;
Erkrankung: Erkrankung aufgetreten

\neg Symptom: Symptom nicht vorhanden
 \neg Erkrankung: Erkrankung nicht aufgetreten

Die Sensitivität beschreibt den Anteil der Personen mit dem Symptom von den Personen mit der Erkrankung bzw. die Wahrscheinlichkeit für das Auftreten eines Symptoms bei einer Person mit der Erkrankung. Die Spezifität ermittelt sich umgekehrt als der Anteil der Personen ohne das Symptom von den nicht erkrankten Personen bzw. als die Wahrscheinlichkeit für das Nicht-Auftreten eines Symptoms bei einer nicht erkrankten Person. Die positive prädiktive Stärke (PPP) ist definiert durch den Anteil der erkrankten Personen von den Personen mit dem Symptom bzw. als die Wahrscheinlichkeit, mit der eine Person mit dem Symptom erkrankt (ist). Entsprechend bestimmt sich die negative prädiktive Stärke (NPP) durch den Anteil der nicht erkrankten Personen an den Personen ohne das Symptom bzw. als Wahrscheinlichkeit nicht zu erkranken (nicht erkrankt zu sein), wenn das Symptom nicht vorliegt.

Bei der Abschätzung der diagnostischen Güte von Testverfahren, die der Zuordnung von Patienten zu bestimmten Erkrankungsbildern bzw. Risikogruppen dienen, wird darüber hinaus den Anteilen falsch-positiv (%F-P) und falsch-negativ beurteilter Patienten (%F-N) besondere Aufmerksamkeit gewidmet. Dabei werden die Mitglieder einer Risikogruppe, die nicht erkranken als Falsch-Positive (engl.: „false positives“; Bell, 1992, p. 372) bezeichnet, die nicht der Risikogruppe zugerechneten, jedoch erkrankenden Personen als Falsch-Negative (engl.: „false negatives“; Bell, 1992, p. 372):

$$\%F-P = 100 \times \Sigma(\neg \text{Erkrankung} \cap \text{Symptom}) / N \quad \text{und}$$

$$\%F-N = 100 \times \Sigma(\text{Erkrankung} \cap \neg \text{Symptom}) / N.$$

Die Prozentsätze falsch-positiver und falsch-negativer Vorhersagen beschreiben - wie die richtig-positiver und richtig-negativer Vorhersagen - hierbei lediglich Zellenbesetzungen in ei-

ner 4-Felder-Tafel und haben damit nicht notwendigerweise einen Einfluß auf die Frage, ob ein Patient erkranken wird (erkrankt ist) oder nicht. Kraemer (1992) schlußfolgerte daraus, daß anhand dieser Kennwerte „... we cannot assume that the diagnosis is correct and the test is wrong; the opposite may occasionally be true“ (ebd., p. 27). Im allgemeinen wird aber eine große Übereinstimmung zwischen Testergebnis und einer klinisch zuverlässigen Diagnose anzeigen, daß der Test - wie schon die Diagnose - ein valider Indikator für das Vorliegen der Erkrankung oder des Erkrankungsrisikos ist, eine geringe Übereinstimmung, also hohe Prozentsätze falscher Vorhersagen, hingegen auf einen mangelhaft validen Test hinweisen.

Die einzelnen diagnostischen Kennwerte sind dabei nicht unabhängig voneinander: So ist etwa ein Symptom dann hoch sensitiv, wenn es bei vielen erkrankten Personen vorkommt. In der Praxis, insbesondere in der Früherkennung initialer schizophrener Prodrome, hat dieses häufige Auftreten aber vielfach zur Folge, daß das entsprechende Symptom auch bei nicht erkrankten Patienten zahlreich anzutreffen ist, was eine erhöhte Rate falsch-positiver Vorhersagen bei gleichzeitiger Minderung der PPP zur Folge hat. Zudem verringert sich die Zahl nicht-erkrankter Personen ohne das Symptom, was zu einem unerwünschten Absinken der Spezifität führt. Die geringe Zahl von Personen ohne das Symptom wird jedoch auch eine Erhöhung der NPP und eine Reduktion der Rate falsch-negativer Vorhersagen mit sich bringen. Umgekehrt wird eher selten auftretendes, aber für ein Prodrom möglicherweise charakteristischeres Symptom nur eine geringe Sensitivität, aber auch eine geringe Rate falsch-positiver Vorhersagen aufweisen. Hingegen wird das seltenere Auftreten in der erkrankten Gruppe zu einer Erhöhung des Prozentsatzes falsch-negativer Vorhersagen und einer Verringerung der NPP führen, das noch seltenere Auftreten in der nicht-erkrankten Gruppe aber zu einer Steigerung der PPP und der Spezifität. Optimale Werte für alle diagnostischen Effizienzindizes werden dann erzielt, wenn das Symptom ausschließlich in der erkrankten Gruppe und hier bei allen Personen auftritt. Die Sensitivität, Spezifität, PPP und NPP besäßen in diesem Fall Prozentwerte von 100% bzw. Wahrscheinlichkeiten von 1, die Raten falscher Vorhersagen betrügen 0%.

Sensitivität und Spezifität sind die am häufigsten verwandten Parameter, da sie im Gegensatz zu PPP und NPP unabhängig von der Prävalenz der Erkrankung in der Stichprobe interpretiert werden können und darüber hinaus auch über Fallstudien mit Kontrollgruppen geschätzt werden können, die erheblich kostengünstiger als Kohortenstudien sind (Leisering & Pepe, 1998). Problematisch wirkt sich jedoch ein möglicher Bias bei der Stichprobenauswahl (Überweisungs-Bias), bei der möglicherweise willkürlichen Dichotomisierung der abhängigen Variablen 'Erkrankung' und/oder bei der Auswahl der nachzuuntersuchenden Fälle (Verifizierungs-Bias) aus (Begg & Greenes, 1983).

Im Hinblick auf die Auswahl eines geeigneten diagnostischen Testverfahrens wurden mehrere Daumenregeln (engl.: „rules of thumb“; Boyko, 1994, p. 175) vorgeschlagen: Nach Sackett (1991, 1992) solle beispielsweise der spezifischste Test für die Diagnosestellung der entsprechenden Erkrankung herangezogen werden, der sensitivste Test hingegen für deren Ausschluß, ohne dabei die Anzahl falsch-positiver und falsch-negativer Klassifikationen zu berücksichtigen. McNeil, Keeler und Adelstein (1975) schlugen demgegenüber eine Auswahlstrategie vor, die sich an den relativen praktischen Kosten von Fehlklassifikationen orientiert. Hiernach solle ein diagnostischer Test für eine ernste Erkrankung, etwa einer Schilddrüsenunterfunktion, für die es jedoch eine auch bei fälschlich positiv diagnostizierten Personen gut verträgliche und sicherere Behandlungsmethode gebe, möglichst hoch sensitiv sein, auch wenn dies zu Lasten der Spezifität gehe. Eine solche Strategie wäre etwa dann für die Früherkennung von schizophrenen Erkrankungen empfehlenswert, wenn die Diagnose eines initialen Prodroms einer Schizophrenie nur eine symptomorientierte Psychotherapie nach sich zöge. Im Gegensatz dazu forderten die Autoren (ebd.) eine möglichst hohe Spezifität auch auf Kosten einer geringen Sensitivität für die Diagnose von Erkrankungen, deren Behandlung mit starken Nebenwirkungen und nicht abzuschätzenden Spätfolgen einher gehe, etwa der eines Non-Hodgkin-Lymphoms oder einer neuroleptischen Medikation als Konsequenz der Diagnose eines schizophrenen Prodroms. Andere Testauswahlstrategien zielen auf eine Nutzen-Maximierung bzw. Erfassung der Erkrankungswahrscheinlichkeit ab, da sich zeigte, daß eine Auswahl nach Sacketts Daumenregel (1991, 1992), also allein auf Basis der Spezifität, im Einzelfall eine zu geringe PPP besitzt (Corpuz, Hilton, Lardis, Singer & Zolan, 1991; Newman et al., 1991).

Auf Einzelsymptomebene sind Sensitivität und Spezifität eines Merkmals heute noch weitgehend die ausschlaggebenden Kriterien bei der Beurteilung der Tauglichkeit eines diagnostischen Kriteriums oder diagnostisch bzw. prognostisch relevanten Symptoms. So sahen etwa Flaum und Andreasen (1991) für die Sensitivität eines zur Diagnosestellung einer schizophrenen Störung geeigneten Merkmals Werte unterhalb einer Wahrscheinlichkeit von 0,10 bis 0,15 auch bei einer Spezifität von 1 als ungenügend an, da ein solches Merkmal zu selten vorkomme, um eine verlässliche Beurteilung im Einzelfall zu erlauben, nämlich nur bei 10 bis 15% der Erkrankten. Hingegen wurden Sensitivitätswerte über 0,30 bis 0,40 bei einer nicht näher quantifizierten guten Spezifität von ihnen als optimal für diagnostisch relevante Symptome bei einer in ihrem klinischen Bild variierenden Störung wie der Schizophrenie eingestuft (ebd.).

Neben der Sensitivität ist in der Früherkennung die PPP, die Genauigkeit eines Tests oder Symptoms bei der Vorhersage der Erkrankung, ein wichtiges Kriterium, denn ein diagnostisch relevantes Prodromalsymptom sollte nicht nur häufig genug auftreten, um die Diagno-

se eines Prodroms zu erlauben, sondern auch den späteren Krankheitsausbruch mit hinlänglicher Sicherheit präzisieren können. Hierbei werden auf Einzelitemebene Wahrscheinlichkeiten von mindestens 0,70 bereits als hoch angesehen (Jackson, McGorry & Dudgeon, 1995). Da jedoch PPP und NPP von der Prävalenz der Erkrankung in der Stichprobe abhängig sind, ist ihre Generalisierbarkeit nicht generell gewährleistet. Eine Übertragbarkeit auf nicht zu der Berechnungsstichprobe gehörige Fälle ist somit streng genommen nur dann möglich, wenn die Werte an einer repräsentativen Stichprobe etwa aus der Normalbevölkerung gewonnen wurden.

Eine ebenfalls wichtige Rolle, insbesondere im Hinblick auf medizinisch-ethische Aspekte der Früherkennung und Prävention, spielen die falsch-positiven Vorhersagen, da bei einer großen Zahl falsch-positiv beurteilter Patienten der Einsatz präventiver Interventionen im Einzelfall nicht mehr gerechtfertigt sein kann, „...because a majority of cases considered at risk could suffer negative effects of intervention along with positive effects that they do not need.“ (Bell, 1992, p. 380). Hierbei sind die Kosten einer möglicherweise nicht indizierten Behandlung gegen die einer versäumten Frühbehandlung nicht erkannter Fälle, der falsch-negativ beurteilten Patienten, abzuwägen (Boyko, 1994; Weinstein et al., 1980), wie dies ja schon mit der ‘Daumenregel’ von McNeil, Keeler und Adelstein (1975) angestrebt wurde. Verbindliche Auswahlkriterien, die alle Kennwerte berücksichtigen und gegeneinander abwägen, existieren jedoch zur Zeit nicht.

2.2.5.1 Diagnostische Likelihood Ratios

Aufgrund der Problematik bei der Verwendung der ‘klassischen’ diagnostischen Effizienzindizes wurde alternativ vorgeschlagen, die positive (+) und negative (–) Vorhersagekraft eines Tests über die entsprechenden diagnostischen Likelihood Ratios (DLR) zu bestimmen, die Sensitivität und Spezifität gegeneinander abwägen und wie folgt definiert sind (Leisenring & Pepe, 1998; Sackett, Haynes & Tugwell, 1985):

$$\begin{aligned} \text{DLR}^+ &= \text{Sensitivität} / (1\text{-Spezifität}) \\ &= p(\text{Symptom} \mid \text{Erkrankung}) / p(\text{Symptom} \mid \neg \text{Erkrankung}) \end{aligned}$$

$$\begin{aligned} \text{DLR}^- &= (1\text{-Sensitivität}) / \text{Spezifität} \\ &= p(\neg \text{Symptom} \mid \text{Erkrankung}) / p(\neg \text{Symptom} \mid \neg \text{Erkrankung}), \end{aligned}$$

wobei Symptom: Symptom vorhanden; – Symptom: Symptom nicht vorhanden,
 Erkrankung: Erkrankung aufgetreten – Erkrankung: Erkrankung nicht aufgetreten.

Eine $\text{DLR}^+ > 1$ kennzeichnet einen Test mit einer besseren Einschlußfähigkeit der Erkrankung, wobei der Wert den Faktor angibt, um den die Erkrankungswahrscheinlichkeit bei einem

positiven Testergebnis steigt. Hingegen zeigt eine $DLR^- < 1$ einen Test mit besserer Ausschlußfähigkeit an, d.h. ein Patient mit einem negativen Testresultat hat eine um $(1 - DLR^-)$ Prozent geringere Erkrankungswahrscheinlichkeit im Vergleich zu seiner Erkrankungswahrscheinlichkeit vor Kenntnis des Testresultats (Boyko, 1994; Leisenring & Pepe, 1998). Bei einem Wert von 1 für die DLRs resultiert aus der Kenntnis des Testergebnisses keine zusätzliche Information über die Erkrankungswahrscheinlichkeit (Leisenring & Pepe, 1998). Nach Ansicht der Autoren kann die Anwendung von DLRs „...the common misconception that low sensibility translates to low ability to rule in disease and high specificity to increased ability to rule out disease“ (ebd., p. 446) entgegenwirken, da in der Tat das Gegenteil zutreffend sein kann. So würde bei einem hypothetischen diagnostischen Test, der eine Sensitivität von 0,20 und eine Spezifität von 0,90 aufweist, die $DLR^+ = 2,00$ und die $DLR^- = 0,90$ betragen, d.h. bei einem positiven Testergebnis wäre die Erkrankungswahrscheinlichkeit verdoppelt, bei einem negativen hingegen nur um 10% reduziert.

Wie Sensitivität und Spezifität haben auch die diagnostischen Likelihood Ratios als Maß für die Fähigkeit eines diagnostischen Tests, eine Erkrankung ein- oder auszuschließen, den Vorteil, daß sie von der Prävalenz der Erkrankung in der Untersuchungsstichprobe unabhängig sind (ebd.). Die Entscheidung zwischen zwei Tests bzw. zwei Symptomen aufgrund von Sensitivität und Spezifität ist nur dann eindeutig, wenn die Werte in einem der beiden Parameter in beiden Tests identisch sind. Dabei wird der sensitivere von zwei Tests zum Ausschluß der Erkrankung bei gleicher oder höherer Spezifität im Vergleich zum Alternativtest die Erkrankung mit größerer Sicherheit ausschließen, während für die Diagnosestellung der spezifischere von zwei Tests bei vergleichbarer oder höherer Sensitivität eine größere Verlässlichkeit der Diagnose gewährleisten wird. Sind jedoch die Werte für Sensitivität und Spezifität in beiden Tests unterschiedlich, geben DLRs eine bessere Abschätzung der diagnostischen Güte eines Tests als die alleinige Betrachtung von Sensitivität und Spezifität (Boyko, 1994; Schultze-Lutter, 2000). Allerdings kann es gerade auf Einzelsymptomebene bei der Berechnung von DLR^+ zu Schwierigkeiten kommen, wenn das Symptom in der erkrankten Gruppe niemals aufgetreten ist, da dies zu einer unzulässigen Division durch Null führt. In diesem Fall ist es aber ohnehin sehr fraglich, ob das Symptom überhaupt einen Beitrag zur Identifizierung der Erkrankung leisten kann.

Leisenring und Pepe (1998) betrachten die DLRs als alternative oder ergänzende Meßmethode zu den Standardmessungen Sensitivität und Spezifität, mit denen sie den Vorzug der funktionellen Unabhängigkeit von der Prävalenz der Erkrankung in der Stichprobe gemeinsam haben, aber im Gegensatz zu letzteren einen direkten Bezug zu dem Prädiktionwert des Tests besitzen und somit für den Kliniker leichter zu interpretieren seien. Dennoch fand die Berechnung von diagnostischen Likelihood Ratios auf Einzelsymptomebene für die Be-

stimmung der Erkrankungswahrscheinlichkeit bei Vorliegen bzw. Nicht-Vorliegen eines Symptoms bislang - zumindest in der psychiatrischen Literatur, wie überhaupt in der biostatistischen Literatur (ebd.) - noch keine bzw. nur geringe Berücksichtigung.

2.2.6 Vergleichsstudien

Obwohl in den letzten Jahrzehnten immer größere Anstrengungen unternommen wurden, für Nominaldaten Techniken und Strategien zur Generierung von Vorhersagemodellen mit und ohne Option zur Selektion von Variablen neu zu entwickeln oder aber bestehende auf dieses Skalenniveau auszudehnen, widmeten sich nur wenige Studien einem Vergleich der Vorhersagequalität verschiedener Methoden. Ein umfassender Vergleich von acht Methoden zur Prognose eines dichotomen Ausgangs wurde 1994 von Marshall, Grover, Henderson und Hammermeister vorgelegt. Hierin wurden anhand von 33 unabhängigen Variablen unterschiedlichen Skalenniveaus, die an insgesamt 12 712 Patienten (Berechnungsstichprobe: N = 6 317 und Validierungsstichprobe: N = 6 395) erhoben worden waren, acht Modelle zur Vorhersage von Todesfällen nach Herzoperationen im Hinblick auf ihre Diskriminationsleistung verglichen. Dies waren im einzelnen:

1. schrittweise logistische Regressionsanalyse der unabhängigen Variablen,
2. logistische Regressionsanalyse aller unabhängiger Variablen,
3. schrittweise logistische Regressionsanalyse mit vorheriger Clusteranalyse der unabhängigen Variablen und Clusteroptimierung nach klinischem Urteil,
4. schrittweise logistische Regressionsanalyse im Anschluß an eine Hauptkomponentenanalyse,
5. subjektiver 'Krankheitswert' („sickness score“: ebd., p. 1506) generiert nach klinischem Urteil,
6. diskriminanzanalytisches Modell unter Einbezug der Apriori-Wahrscheinlichkeiten des Ausgangs basierend auf dem Bayes-Theorem über die auf Einzelitemebene signifikant mit dem Ausgang assoziierten Variablen (χ^2 - oder t-test, $p < 0,01$),
7. additives Modell basierend auf dem Additionstheorem für Wahrscheinlichkeiten unter Verwendung des nicht adjustierten relativen Risikos der Einzelvariablen über die auf Einzelitemebene signifikant mit dem Ausgang assoziierten Variablen (χ^2 - oder t-test, $p < 0,01$) und
8. rekursive Partitionierung (CART; Breiman et al., 1984).

Für die Bestimmung der Güte der Vorhersagegenauigkeit waren von Harrell, Califf, Pryor, Lee und Rosati (1982) zwei Maße vorgeschlagen worden: Reliabilität und Diskriminationsfähigkeit. Die Reliabilität bestimmt sich durch die Übereinstimmung zwischen vorhergesagtem

und beobachtetem Ausgang; die Diskriminationsfähigkeit bezieht sich auf das Ausmaß, mit dem die beiden möglichen Ausgänge der abhängigen Variablen unterschieden werden können. Für die Berechnung der Diskriminationsfähigkeit empfehlen sie den Konkordanz-Index c , der im Falle einer binären abhängigen Variablen der Fläche unter der 'Receiver Operating Characteristic' (ROC) Kurve entspricht (Hanley & McNeill, 1982). Ein $c = 0,5$ entspräche hierbei der Zufallsfunktion mit gleicher Wahrscheinlichkeit für das Auftreten einer der beiden Ausgänge, ein $c = 1,0$ kennzeichnete ein Instrument mit perfekter Diskriminationsfähigkeit, das alle Objekte entsprechend ihrer beobachteten Ausprägung auf der abhängigen Variablen korrekt zuordnen würde.

Marshall und Kollegen (1994) verwendeten die Diskriminationsfähigkeit für den Vergleich der Vorhersageleistung der acht generierten Modelle. Die schrittweise logistische Regression über alle Kandidatorvariablen und die schrittweise logistische Regression über zuvor berechnete Cluster zeigten hierbei die besten Vorhersagestärken sowohl für die Berechnungs- als auch für die Validierungsstichprobe (Tabelle 8). Wie zu erwarten, war die Fläche unter der ROC-Kurve, der Konkordanz-Index c , in der Validierungsstichprobe generell kleiner als die der Berechnungsstichprobe (Tab. 8), was darauf hindeutete, daß alle Verfahren - mit Ausnahme des nach dem klinischen Urteil gebildeten Krankheitswerts - zu einer Überanpassung der Vorhersagemodelle an die empirischen Daten tendierten, und erneut bestätigte, daß eine Validierung der Ergebnisse anhand einer unabhängigen Stichprobe unerlässlich für die Beurteilung einer späteren Prognose im Einzelfall ist.

Tabelle 8: Ergebnisse der Beurteilung verschiedener Vorhersagemodelle in absteigender Ordnung: Anzahl der im Modell enthaltenen Parameter, Diskriminationsfähigkeit gemessen an dem Konkordanz-Index c für Berechnungs- und Validierungsstichprobe, Prozentsatz der aus der Berechnung wegen unvollständiger Daten ausgeschlossenen Beobachtungen und Standardfehler von c in der Validierungsstichprobe (nach Marshall et al., 1994).

Modell	Parameteranzahl	Berechnungsstichprobe		Validierungsstichprobe	
		c Index	% ausgeschlossen	c Index	Standardfehler
Clusteranalyse (3)	6	0,733	19	0,711	0,019
schrittweise (1)	6	0,739	13	0,710	0,019
additives Modell (7)	24	0,718	20	0,697	0,021
Bayes Modell (6)	24	0,713	20	0,695	0,021
logistisch (2)	33	0,749	38	0,694	0,022
Hauptkomponenten (4)	5	0,700	38	0,690	0,022
Krankheitswert (5)	8	0,674	24	0,678	0,020
Partitionierung (8)	11	0,716	0	0,655	0,017

Die in Klammern gesetzten Numerierungen hinter den Modellen beziehen sich auf die Auflistung der Methoden (s.o.).

Eine teilweise Stützung dieser Ergebnisse stellt die Untersuchung von Harrell, Lee, Matchar und Reichert (1985) dar. Sie verglichen schrittweise logistische Regressionsanalysen über alle unabhängigen Variablen und über einen auf Clusteranalysen der Variablen basierenden Krankheitswert bzw. zuvor berechnete Hauptkomponenten mit der rekursiven Partitionierung (CART; Breiman et al., 1984) und einem zweiten Stratifizierungsverfahren, der Entropie-Minimax-Methode (Christensen, 1981), anhand von 25 unabhängigen Variablen unterschiedlichen Skalenniveaus zur Vorhersage des Behandlungsausgangs (vollständige versus unvollständige Remission) von insgesamt 630 Patienten mit Hodgkin-Lymphom (Berechnungsstichprobe 1: N = 110; Berechnungsstichprobe 2: N = 224; Validierungsstichprobe: N = 296). Die regressionsanalytischen Modelle zeigten sich mit Ausnahme der schrittweisen logistischen Regression über alle Kandidatorvariablen im Falle der kleinen Berechnungsstichprobe der rekursiven Partitionierung generell überlegen, wobei die Regressionsmodelle der Hauptkomponenten über beide Berechnungsstichproben die besten Validierungsergebnisse gemessen an der Diskriminationsfähigkeit erlangten. Im Falle der Berechnung über die größere Stichprobe zeigte die Entropie-Minimax-Methode jedoch eine gleich hohe Diskriminationsstärke wie die logistische Regression über die Hauptkomponenten.

Eine bessere Klassifizierungsleistung der schrittweisen logistischen Regressionsanalyse im Vergleich zur hierarchischen Partitionierung (CART; Breiman et al., 1984) fanden auch Knuiman, Vu und Segal (1997). Ihre Stichprobe bestand aus 1 701 Personen, die auf acht potentielle Risikofaktoren für kardiovaskuläre Erkrankungen untersucht wurden, wie etwa Raucher, Cholesterinspiegel und Blutdruckwerte. Die abhängige Variable war Tod durch eine kardiovaskuläre Erkrankung im Zeitraum des 13-Jahres-Follow-up. Dieses Ereignis trat bei 137 Personen ein. In der schrittweisen logistischen Regression wurden neben den Einzelvariablen auch alle paarweisen Interaktionen mit berücksichtigt, und in Ermangelung einer Validierungsstichprobe wurde eine 10-fache Kreuzvalidierung der Ergebnisse vorgenommen. Während das logistische Regressionsmodell zu einer Fläche unter der ROC-Kurve von $c = 0,72$ führte, betrug diese für das Partitionsmodell nur $c = 0,66$. Eine ebenfalls leichte Überlegenheit der logistischen Regression gegenüber der hierarchischen Partitionierung zeigten auch drei weitere Studien (Hadorn, Draper, Rogers, Keeler & Brook, 1992; Long, Griffith, Selker & D'Agostino, 1993; Selker, Griffith, Patil, Long & D'Agostino, 1995). Dennoch sehen Knuiman, Vu und Segal (1997) als Vorteil der hierarchischen Partitionierung, daß sie keine Vorannahmen über das Skalenniveau oder den Trend auf dieser Skala macht und somit auch nicht-lineare Beziehungen zwischen der abhängigen und den unabhängigen Variablen zuläßt (Marshall et al., 1994), daß sie in der Lage ist, Interaktionen und Synergismen zwischen Prädiktoren aufzudecken und eine mögliche unterschiedliche Wichtigkeit einzelner Risikofaktoren in verschiedenen Risikogruppen berücksichtigen kann. Deshalb kommen sie zu

dem Schluß, daß „...it is not clear on *a priori* grounds whether [...] logistic [...] regression would necessarily have superior performance in risk estimation and prediction compared with tree approaches.“ (Knuiman, Vu & Segal, 1997, p. 132).

So fanden im Gegensatz zu diesen Ergebnissen eine Reihe anderer Studien zum Vergleich der Vorhersagequalität von logistischen Regressionsanalysen und Partitionierungsverfahren auch in der Tat zumindest eine vergleichbare Vorhersagestärke der beiden Ansätze (Cook & Goldman, 1984; Marshall, 1986; Marshall & Chisholm, 1985, van Mechelen & DeBoeck, 1990).

Van Mechelen und DeBoeck (1990) verglichen die logistische Regression und die hierarchische Partitionierung (HICLAS; DeBoeck & Rosenberg, 1988) anhand der Beurteilung von 23 Symptomen bei 30 Patienten durch 15 Psychiater hinsichtlich der kategorialen Zielvariable 'Diagnose' = {affektive Störung; Schizophrenie; Angststörung} und kamen zu ähnlichen Ergebnissen für beide Verfahren nicht nur in Bezug auf die Klassifizierungsleistung, sondern auch auf die Variablenauswahl. Hierbei unternahm die logistische Regression teils weitere Schritte zur Generierung eines auf Kriteriumwahrscheinlichkeiten basierten feineren Modells als der Alles-oder-Nichts-Ansatz des Partitionierungsmodells. Allerdings verbargen sich die Interaktionen der unabhängigen Variablen hinter dem Muster der Regressionskoeffizienten und waren dadurch - anders als bei den mit der Partitionierungsanalyse ermittelten Bool-schen Entscheidungsregeln - schwerer zu interpretieren. Die Autoren zogen daher den Schluß: „...the logical and the logistic approaches can be considered complementary in that the former leads to easily interpretable categorical combination rules, whereas the latter provides a detailed model of criterion probabilities.“ (ebd., p. 693).

Ausgehend von 50 unabhängigen Variablen verglichen Cook und Goldman (1984) die logistische Regressionsanalyse, die lineare Diskriminanzanalyse und die rekursive Partitionierung (CART; Breiman et al., 1984) über die Sensitivitäten und Spezifitäten für unterschiedliche Schwellenwerte zur Vorhersage eines myokardialen Infarkttrisikos bei Patienten mit Brustschmerzen in der Notfallaufnahme (N = 482) anhand der Fläche unter den resultierenden ROC-Kurven. Sie fanden vergleichbare Vorhersageleistungen beider Verfahren in einer Validierungsstichprobe (N = 900) bis zu einer Sensitivität von 85%, bei höheren Sensitivitätswerten aber signifikant höhere Werte der Spezifität für das anhand der rekursiven Partitionierung generierte Modell. In der Diskussion ihrer Ergebnisse kamen sie dennoch zu dem Schluß, daß keine der drei Analysemethoden, logistische Regression, lineare Diskriminanzanalyse und rekursive Partitionierung, eindeutig die überlegenere sei, sondern daß jede dieser Methoden Eigenschaften besäße, die ihre Anwendung je nach Situation wünschenswert mache. Die Autoren sahen die folgenden neun Charakteristiken der rekursiven Partitionierung

als Vorzüge bzw. Nachteile gegenüber konventionelleren multivariaten Verfahren wie Regressions- und Diskriminanzanalysen an:

1. Die rekursive Partitionierung ist eine einfache und verständliche Methode zur Klassifikation von Objekten: Der in konventionellen multivariaten Verfahren berechnete Summenwert der Produkte 'unabhängige Variable \times Gewichtung', der mit einem Apriori-Schwellenwert verglichen wird, um den Ausgang vorherzusagen, ist teils nicht nur schwer zu interpretieren, sondern kann auch in seiner dahinterstehenden Berechnung kompliziert sein.
2. Die rekursive Partitionierung basiert auf der Identifizierung von Interaktionen zwischen den unabhängigen Variablen: Nur in wenigen Fällen werden in konventionellen multivariaten Modellen Interaktionen der Variablen berücksichtigt, die die Art, in der die Diskriminationsstärke einer Variable durch die An- oder Abwesenheit einer anderen beeinflusst wird, abbilden. Die Vernachlässigung solcher Interaktionen erfolgt meist entweder aus Gründen der Vermeidung zu abstrakter Modelle oder weil die genaue Art positiver und negativer Interaktionen - insbesondere bei einer großen Anzahl von unabhängigen Variablen - unbekannt ist.
3. Die rekursive Partitionierung beruht nicht auf einer möglicherweise unzutreffenden Linearitätsannahme und kann nicht-lineare Beziehungen unabhängiger und abhängiger Variablen identifizieren.
4. Die rekursive Partitionierung bietet einfache Möglichkeiten zur Gruppierung von Objekten in homogene Risikoschichten (engl.: „risk strata“) mit unterschiedlichen Endpunktwahrscheinlichkeiten oder zur detaillierten Bildung von Subgruppen, die einer weiteren Analyse unterzogen werden können, etwa im Hinblick auf die adäquateste Behandlungsmethode für einzelne Gruppen.
5. Die rekursive Partitionierung berücksichtigt zuvor festgelegte Wahrscheinlichkeiten und Straffunktionen für Fehlklassifizierungen bei der Variablenselektion. Im Gegensatz zu konventionellen multivariaten Verfahren, die die Kosten falsch-positiver Vorhersagen denen falsch-negativer gleichsetzen, werden a priori festgelegte Wahrscheinlichkeiten und Straffunktionen für Fehlklassifizierungen bei der Variablenauswahl der rekursiven Partitionierung berücksichtigt.
6. Die rekursive Partitionierung kann möglicherweise zusätzliche prädiktive Variablen in späteren Phasen des Selektionsprozesses übergehen, da die den Berechnungen zugrunde liegenden Stichproben im Verlauf der Analyse immer kleiner werden und somit die Aufdeckung kleinerer Effekte kaum mehr möglich ist.
7. Die rekursive Partitionierung verstärkt möglicherweise das Problem multipler Testungen: Bei der Anwendung von konventionellen multivariaten Verfahren bleibt das Problem der Fehlerkumulierung durch multiple Einzeltestungen meist unberücksichtigt. Dies Problem

verstärkt sich in der rekursiven Partitionierung, bei der in jedem Schritt eine große Zahl möglicher Partitionen berücksichtigt werden. Da darüber hinaus nur binäre Partitionen einbezogen werden, resultiert schon aus dem Einbezug einer ordinalen Variable eine Vielzahl möglicher binärer Partitionen.

8. Die rekursive Partitionierung verstärkt möglicherweise das Problem der Überanpassung, da die Endpunkte nur noch auf einer geringen Stichprobengröße basieren und somit unter Umständen nur ein Artefakt ihrer Stichprobe darstellen und somit nicht generalisierbar sind. Eine Überprüfung anhand einer Validierungsstichprobe ist daher unerlässlich.
9. Die rekursive Partitionierung unterschätzt womöglich die prädiktive Stärke intervallskalierter unabhängiger Variablen, die für die Berechnung künstlich in Intervalle zerlegt werden und somit einen Teil ihrer Information verlieren.

Cook und Goldman (1984) forderten in ihrer Schlußfolgerung trotz offensichtlicher Nachteile auch auf Seiten der rekursiver Partitionierung ihren verstärkten Einsatz, insbesondere bei Vorliegen größerer Stichproben und/oder des Verdachts, daß die Vorhersagefähigkeit einer unabhängigen Variable möglicherweise über Kategorien anderer unabhängiger Variablen variiert. In Übereinstimmung mit van Mechelen und DeBoeck (1990) empfehlen sie, daß „Even in situations where a parametric analysis [logistische Regressions- und lineare Diskriminanzanalyse werden von den Autoren aufgrund der in den postulierten Beziehungen involvierten unbekanntem Parameter für die prädiktiven Variablen als parametrische Analysen bezeichnet, v. Verf.] ultimately may be preferable, recursive partitioning analysis should be considered to assist in the identification of interaction terms that may then by [*sic*] used in the construction of a parametric model.“ (Cook & Goldman, 1984, p. 730).

Ein Vergleich von logistischer Regression und einer ersten Version der 'Search Partition Analysis - SPAN' (Marshall, 1995) anhand von 22 Merkmalen zur Diagnose von gastrointestinalem Krebs, die bei 195 Personen erhoben wurden, ergab nur marginale Unterschiede der beiden Verfahren dahingehend, daß die Vorhersageleistung des Regressionsmodells besser für den Ausgang 'Krebs' (PPP = 0,96 im Gegensatz zu 0,74 bei SPAN), die der Partitionierungsanalyse besser für den Ausgang 'benigner Krebs' (NPP = 0,96 im Gegensatz zu 0,76 für das Regressionsmodell) war (Marshall, 1986; Marshall & Chisholm, 1985).

Zwei Vergleiche von hierarchischer (S routines; Clark & Pregibon, 1992) und nicht-hierarchischer Partitionierung (SPAN) zum einen zur Vorhersage eines niedrigen Geburtsgewichts anhand von acht Variablen (N=188) und zum anderen zur Vorhersage von Glukoseintoleranz (N=5 314) anhand einer nicht näher spezifizierten Anzahl von mehr als sieben Variablen, zeigte bei minimalsten Unterschieden in der Diskriminationsstärke beider Verfahren eine leichte Überlegenheit der nicht-hierarchischen Methode, SPAN, im Hinblick auf die

positive prädiktive Stärke im zweiten Vergleich (S routines: PPP = 0,135 vs. SPAN: PPP = 0,144) bei gleicher negativer prädiktiver Stärke (NPP = 0,984) beider Vorhersagemodelle (Marshall, 1995).

1993 unternahmen Hashemi, Jelovsek und Razzaghi einen Vergleich zwischen zwei Grob-Mengen-theoretisch basierten Ansätzen und einer Diskriminanzanalyse (s. Tab. 9). Während ein Grob-Mengen-Modell eng an die ursprüngliche Konzeption von Pawlak (1982, 1984) angelehnt ist, wird in dem modifizierten Ansatz durch Schätzung der Regeln ein Einbezug der Grenzregion in die Entscheidungsfindung möglich, ähnlich wie dies durch die Einführung von 'rule roughness' und 'rule precision threshold' auch bei der neuen Version von DataLogic (vgl. Kap. 2.2.5) ermöglicht ist.

Tab. 9: Klassifizierungsraten für einen Grob-Mengen-Ansatz und einen modifizierten Grob-Mengen-Ansatz (geschätzte Regeln) im Vergleich zur Diskriminanzanalyse für a) die Toxizitätsdaten und b) die zufällig generierten Daten (nach Hashemi, Jelovsek & Razzaghi, 1993).

Modell	Analyse- bedingung (Kategorien der Zielvariablen) ¹	Gesamtge- nauigkeit (%) ²	Diagnostische Genauigkeit		Prädiktive Genauigkeit	
			Positive Fälle ³	Negative Fälle ⁴	Positiv ⁵	Negativ ⁶
a)						
Diskriminanzanalyse	[+][−] N=116	86,2	56,3	97,6	90	85,4
Grob-Mengen	[+][−] N=116	73,3	78,1	71,4	***	***
mod. Grob-Mengen	[+][−] N=116	92,2	84,4	95,2	96,4	97,6
Diskriminanzanalyse	[+][0][?][−] N=175	61,1	40,6	89,3	76,5	62
Grob-Mengen	[+][0][?][−] N=175	59,4	71,9	58,3	***	***
mod. Grob-Mengen	[+][0][?][−] N=175	79,4	78,1	84,5	96,1	92,2
b)						
Diskriminanzanalyse	[+][−] N=116	89,7	65,4	96,7	96,7	90,6
Grob-Mengen	[+][−] N=116	76,7	80,8	73,3	***	***
mod. Grob-Mengen	[+][−] N=116	97,4	88,5	100	100	100
Diskriminanzanalyse	[+][0][?][−] N=175	60	34,5	89,2	89,2	61,2
Grob-Mengen	[+][0][?][−] N=175	60,6	69	61,4	***	***
mod. Grob-Mengen	[+][0][?][−] N=175	79,4	82,8	86,7	92,3	92,3

¹ [−]=negativ, [0]=unbekannt, [?]=Verdacht, [+]=positiv

² Prozentsatz richtig positiver und richtig negativer Vorhersagen von der Gesamtmenge

³ Prozentsatz der richtig positiven Vorhersagen von den Positiven (Sensitivität)

⁴ Prozentsatz der richtig negativen Vorhersagen von den Negativen (Spezifität)

⁵ Prozentsatz der Positiven von den positiven Vorhersagen

⁶ Prozentsatz der Negativen von den negativen Vorhersagen

*** Berechnung nicht möglich

Die Daten bestanden aus 175 Substanzen, die in zwei Analysereihen anhand von 16 unabhängigen Variablen entsprechend ihrer Karzinogenität und Mutagenität ('negativ', 'nicht ge-

testet', 'widersprüchliche Testresultate', 'positiv' als Ausprägungen der abhängigen Variable über ein $N=175$ bzw. nur Berücksichtigung der Ausprägungen 'negativ' und 'positiv' über ein $N=116$) zu klassifizieren waren. Die Vorhersagemodelle wurden anschließend kreuzvalidiert, da eine Validierungsstichprobe nicht zur Verfügung stand.

Der modifizierte Grob-Mengen-Ansatz zeigte sich deutlich dem diskriminanzanalytischen überlegen, während der an die ursprüngliche Grob-Mengen-Konzeptualisierung angelehnte Ansatz ohne den expliziten Einbezug der Grenzregion B_{n_A} zu den ungünstigsten Klassifizierungsleistungen führte (s. Tab. 9).

1994 legten Lette und Mitarbeiter einen Vergleich von schrittweiser logistischer Regressionsanalyse und einem artifiziellem neuronalen Netzwerk-Modell (ANN) bei der Vorhersage von kardiologischen Komplikationen bei nicht-kardiologischen operativen Eingriffen vor.

Artifizielle neuronale Netzwerke sind Computerprogramme zur Simulation der Struktur und Funktion eines rudimentären Gehirns. Das grundlegende Verarbeitungselement des ANN ist das in Analogie zum Gehirn benannte *Neuron*, das ein Eingangssignal, einen *Input*, empfängt und transformiert und dieses transformierte Signal als Ausgangssignal, *Output*, zu nachgeschalteten Neuronen weiterleitet. Die grundlegende Netzwerkstruktur besteht aus drei Neuronenlayern: einem Inputlayer, bei dem jedes Neuron einer unabhängigen Variablen entspricht, einem Hidden-Layer, der in etwa die doppelte Anzahl von Neuronen wie der Inputlayer besitzt, und einem Outputlayer, der die Ergebnisklassen der abhängige Variable repräsentiert (Abb. 24).

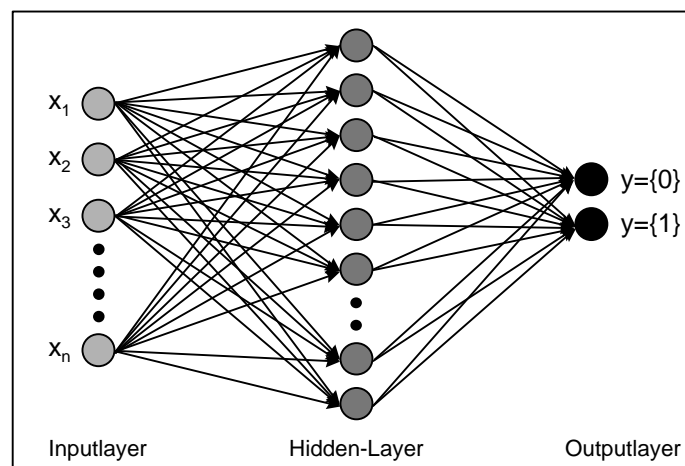


Abb. 24: Vereinfachtes Schema eines artifiziellem neuronalen Netzwerks mit seinen drei Layern.

Jedes Neuron des Inputlayers ist mit jedem Neuron des Hidden-Layers und das wiederum mit den Neuronen des Outputlayers verbunden. Jede Verbindung ist durch ein numerisches Gewicht und seine Richtung charakterisiert und reflektiert damit das Ausmaß des Einflusses eines Neurons auf ein anderes. Dem Inputlayer werden Lernbeispiele (etwa Patientendaten) mehrfach, also iterativ, präsentiert. Dabei kommt es innerhalb des Netzwerks durch Veränderungen der Gewichte der Verbindungen zur Ausbildung einer inneren Repräsentation der im Datensatz enthaltenen Muster. Durch die Neugewichtung der Neuronenverbindungen zwischen Input- und Hidden-Layer werden so lineare und/oder non-lineare Korrelationen zwischen den Ergebnisklassen der abhängigen Variablen und den einzelnen unabhängigen Variablen bzw. Interaktionen von unabhängigen Variablen dargestellt. Nach Abschluß der Lernphase, d.h. nach Erreichen einer Netzwerkkonvergenz, die dadurch gekennzeichnet ist, daß die Signale der Neurone des Inputlayers ein korrektes Outputsignal für jeden Patienten der Berechnungsstichprobe erzeugen, ist das ANN in der Lage, gleiche oder ähnliche Inputmuster in einem neuen Datensatz wiederzuerkennen.

Als Vorteile neuronaler Netzwerk-Modelle gegenüber herkömmlichen statistischen Verfahren gelten die parallele Datenverarbeitung, die Möglichkeit zur Aufdeckung komplexer non-linearer Zusammenhänge und zur gleichzeitigen Verarbeitung von Daten unterschiedlichen Skalenniveaus oder unklarer innerer Struktur sowie die Klassifizierung auch von Objekten, deren Variablen nicht genau denen der Berechnungsstichprobe entsprechen. Als nachteilig werden hingegen die fehlende Möglichkeit zur Eliminierung redundanter oder nicht relevanter Information und damit zur Variablenreduktion bzw. -selektion, die sehr hohe Anzahl benötigter 'Lernbeispiele', also die große Stichprobengröße bei komplexen Datensätzen, die schwierige Interpretierbarkeit der Klassifikationsergebnisse und die erschwerte Validierbarkeit von ANN angesehen (Winter, Vortrag auf dem Mainzer AGNP-Arbeitstreffen 'Methoden' am 31.01.1997).

Der zugrunde liegende Datensatz der Arbeit von Lette und Mitarbeitern (1994) enthielt als unabhängige Variablen 17 dem Krankenblatt entnommene präoperative klinische Parameter auf Nominalskalenniveau, etwa Raucher, Diabetiker, vorbestehende bekannte Herzprobleme oder Auffälligkeiten im Ruhe-EKG, sowie sieben quantitative Meßparameter einer Thallium-Szintigraphie des Herzens. Als negativer postoperativer Ausgang galt Tod durch Herzversagen und akuter nicht-tödlicher Multiinfarkt. Die Gesamtstichprobe von $N = 360$ wurde in eine Berechnungsstichprobe bestehend aus den ersten 200 aufgenommenen Patienten mit 21 postoperativen kardiologischen Komplikationen und eine Validierungsstichprobe aus den zuletzt aufgenommenen 160 Patienten mit neun kardiologischen Fällen unterteilt.

Die schrittweise logistische Regressionsanalyse erfolgte nach der backward Maximum Likelihood Ratio-Methode über alle 24 unabhängigen Variablen, von denen fünf in die Regressi-

ongleichung eingeschlossen wurden. Die Modellierung des ANN folgte einem 'Backpropagation'-Algorithmus über zehn zuvor unter dem Aspekt der Vermeidung widersprüchlicher Patienteninformation ausgewählte unabhängige Variablen und vier Kreuz-Produkt-Terme. Beim Backpropagation-Algorithmus wird das Outputergebnis mit dem gewünschten Klassifikationsergebnis nach Präsentation eines Musterbeispiels verglichen und ggf. vom Supervisor als Irrtum registriert. Der Irrtum wird als Fehlersignal an die einzelnen Neurone im ANN zurückgegeben, die daraufhin ihre Gewichte in die gewünschte Richtung korrigieren. Die Vorauswahl der Input-Variablen schien den Autoren angemessen, da eine Variablenselektion bzw. -reduktion mit Hilfe von ANN-Verfahren nicht möglich ist.

Tabelle 10: Vorhersageleistung von schrittweiser logistischer Regression (LR) und artifiziellem neuronalem Netzwerk-Modell (ANN) in der Berechnungs- (N = 200) und Validierungsstichprobe (N = 160; nach Lette et al., 1994).

	Berechnungsstichprobe				Validierungsstichprobe				p *
	LR vorhergesagt		ANN vorhergesagt		LR vorhergesagt		ANN vorhergesagt		
	KKK	KK	KKK	KK	KKK	KK	KKK	KK	
KKK (b)	159	20	179	0	124	27	145	6	
KK (b)	3	18	0	21	3	6	3	6	
Sensitivität	0,86		1,0		0,67		0,67		1,0
Spezifität	0,89		1,0		0,82		0,96		< 0,01
PPP	0,47		1,0		0,18		0,50		0,08 ^a
NPP	0,98		1,0		0,98		0,98		1,0

KKK: keine kardiologischen postoperativen Komplikationen

KK: kardiologische postoperative Komplikationen

PPP: positive prädiktive Stärke

NPP: negative prädiktive Stärke

(b): beobachtet

* Statistische Signifikanz in der Vorhersageleistung von LR und ANN in der Validierungsstichprobe

^a Keine statistische Signifikanz aufgrund der wenigen Fälle mit KK (n = 9; 6%)

Tabelle 10 zeigt die Vorhersageleistungen beider Modelle in der Berechnungs- und Validierungsstichprobe. Das ANN-Modell zeigte sich in der Validierungsstichprobe der schrittweisen logistischen Regression im Hinblick auf die Spezifität signifikant und im Hinblick auf die positive prädiktive Stärke noch deutlich überlegen, während Sensitivität und negative prädiktive Stärke nahezu identisch waren. Bei der Beurteilung der Ergebnisse ist jedoch zu bedenken, daß der Anteil der Patienten mit kardiologischen Komplikationen in beiden Stichproben sehr gering war und die Anzahl der unabhängigen Variablen (24 insgesamt) die Anzahl der Patienten mit kardiologischen Komplikationen (n = 21) in der Berechnungsstichprobe überstieg, so daß die Gruppenwahrscheinlichkeiten bzw. Logits der logistischen Regressionsanalyse in

der Validierungsstichprobe nicht als zuverlässige Schätzung angesehen werden können (ebd.), da logistische Regressionsanalysen zu den zuverlässigsten Schätzungen bei einer balancierten Randverteilung der abhängigen Variable neigen (Harrell et al., 1985)

Ein von Eisenstein und Alemi 1996 vorgelegter Vergleich von logistischer Regressionsanalyse, Bayes'schem Modell und artifiziellem neuronalem Netzwerk-Modell (ANN) jeweils über zwei unterschiedliche Stichprobengrößen (N = 100 und N = 600) und zwei unterschiedliche Variablenanzahlen (drei und elf unabhängige Variablen) sowie unterschiedlicher Handhabung von fehlenden Werten (mit und ohne Zulassung fehlender Werte in der Berechnung) und Variablen-Interaktionen (nicht zulässig und zulässig) zeigte über alle Bedingungen weitgehend ähnliche Vorhersageleistungen der drei Verfahren bei Ausschluß fehlender Werte in einer Validierungsstichprobe, bei Zulassung fehlender Werte in den Berechnungen jedoch eine durchgängige Überlegenheit des Bayes'schen Modells (Tab. 11). Eine Variablenselektion, die ohnehin nur durch die logistische Regressionsanalyse zu leisten gewesen wäre, war hierbei nicht angestrebt worden.

Tabelle 11: Diskriminationsleistung von logistischer Regression (LR), Bayes'schem Modell (BM) und artifiziellem neuronalem Netzwerk-Modell (ANN) über die verschiedenen Versuchsbedingungen; Fläche unter der ROC-Kurve in der Validierungsstichprobe (nach Eisenstein & Alemi, 1996).

Modell	Keine fehlenden Werte				Fehlende Werte			
	n=100		n=600		n=100		n=600	
	3 UV	11 UV	3 UV	11 UV	3 UV	11 UV	3 UV	11 UV
LR ohne Interaktion	0,663	0,645	0,627	0,733	0,500	0,500	0,654	0,689
LR mit Interaktion	0,663	0,663	0,627	0,733	0,500	0,500	0,654	0,689
BM ohne Interaktion	0,656	0,749	0,644	0,735	0,653	0,730	0,657	0,730
BM mit Interaktion	0,654	0,729	0,644	0,732	0,654	0,715	0,657	0,727
ANN mit Interaktion	0,663	0,662	0,650	0,746	0,619	0,568	0,650	0,594
ANN ohne Interaktion	0,631	0,600	0,650	0,692	0,549	0,598	0,557	0,534

UV: unabhängige Variable, höchster Wert je Bedingung in Fettdruck.

Insgesamt wurde in allen Vergleichsstudien die logistische Regression unter den konventionelleren statistischen Verfahren am positivsten beurteilt. Darüber hinaus wurde aber auch die gemessen an der Häufigkeit ihrer Anwendung immer noch unterschätzte Rolle alternativer Verfahren zur Generierung von Vorhersagemodellen deutlich².

² Eine Vergleichsstudie, die neben unterschiedlichen konventionellen statistischen Verfahren auch unterschiedliche alternative Ansätze mit einbezieht, konnte auch bei umfassenden Literaturrecherchen in DIMDI nicht gefunden werden.

Diese mangelnde Berücksichtigung alternativer Methoden zeigte sich auch in dem 1997 veröffentlichten Übersichtsartikel 'CATEGORICAL DATA ANALYSIS IN PUBLIC HEALTH' von Preisser und Koch, in dem ausschließlich konventionellere statistische Verfahren, die Analyse von Kontingenztafeln, logistische Regressionsanalysen, GEE- und WSK-Methoden, erläutert werden. Und auch in einer 1992 veröffentlichten Bibliographie zu Analysemetoden für korrelierte kategoriale Daten widmeten sich die weitaus meisten der über 100 aufgeführten Arbeiten Regressionsmodellen und -schätzungen, wobei ein besonderer Schwerpunkt auf logistischen Modellen mit binärer abhängiger Variable lag (Ashby et al., 1992).

2.3 FRAGESTELLUNG

Obwohl in den letzten Jahren eine Vielzahl von mathematischen Verfahren zur Generierung von (reduzierten) Vorhersagemodellen neu entwickelt oder bestehende Verfahren an Variablen mit Ordinal- oder Nominalskalenniveau adaptiert wurden, liegen zur Zeit keine umfassenden spezielle Vergleiche der Klassifizierungsleistung von Methoden vor, die auch bei dichotomen abhängigen und unabhängigen Variablen angewandt werden können, wie sie auf psychopathologischer Ebene in der Forschung zur Früherkennung schizophrener Psychosen hauptsächlich zu finden sind. Dies gilt insbesondere für den Vergleich zwischen verschiedenen nicht-inferenzstatistischen, alternativen Methoden, die der Erstellung von Vorhersagemodellen bei paralleler Variablenselektion dienen (vgl. Kap. 2.6).

Anhand prospektiv gewonnener dichotomer Daten zu möglichen Prodromalsymptomen einer schizophrenen Erstmanifestation sollen deshalb in der vorliegenden Studie die beiden wichtigsten konventionellen inferenzstatistischen Verfahren zur Generierung von Vorhersagemodellen bei gleichzeitiger Variablenselektion, schrittweise nonparametrische Diskriminanz- und logistische Regressionsanalyse, mit zwei alternativen Verfahren, der Search Partition Analysis und einer Analyse basierend auf dem Grob-Mengen-Modell, sowie zwei Selektionsvarianten basierend auf diagnostischen Gütekriterien im Hinblick auf ihre Klassifikationsleistung in einer Validierungsstichprobe verglichen werden. Da hier nur Verfahren verglichen werden sollen, die neben der Generierung eines Vorhersagemodells auch eine Variablenreduktion erlauben, werden reine Klassifikationsmodelle, wie etwa auf dem Bayes'schem Theorem oder dem Additionstheorem für Wahrscheinlichkeiten beruhende und artifizielle neuronale Netzwerk-Modelle, nicht berücksichtigt, da sie keine Option zur Variablenselektion beinhalten. Aufgrund der bisher publizierten Vergleiche ist dabei zu erwarten, daß die logistische Regressionsanalyse der nonparametrischen Diskriminanzanalyse überlegen sein wird, wobei die logistische Regression über zuvor generierte Cluster noch einmal zu besseren Ergebnis-

sen als die über Einzelvariablen führen sollte. Die beiden alternativen Verfahren sollten jeweils zu Klassifikationsleistungen führen, die denen der logistischen Regressionsanalysen vergleichbar sind, und die anhand der diagnostischen Likelihood Ratios selektierten Variablen der Variablenselektion anhand klassischer diagnostischer Effizienzindizes überlegen sein.

Bei der Bewertung der einzelnen Methoden soll jedoch nicht nur der Klassifizierungsleistung der einzelnen Vorhersagemodelle in einer Validierungsstichprobe, sondern darüber hinaus auch erstmals gezielt, wenn auch nur deskriptiv der inhaltlichen Plausibilität und praktischen Anwendbarkeit der generierten Modelle Rechnung getragen werden. Dies erscheint vor allem aus praktischen Erwägungen sinnvoll, da in der klinischen Praxis nur solche Vorhersagemodelle auf Akzeptanz stoßen werden, die für den Anwender aufgrund seiner klinischen Vorerfahrung inhaltlich einsichtig und nachvollziehbar sind. Dies gilt insbesondere für ein solch umstrittenes Gebiet wie der Früherkennung von schizophrenen Psychosen und der damit angestrebten Frühintervention (Heinimaa, 2000).

3 METHODE

3.1 VERSUCHSPLANUNG (1)

3.1.1 Versuchsplan-Anlage (1)

In der vorliegenden Studie zum Vergleich verschiedener Vorgehensweisen zur Generierung von Vorhersagemodellen stellen die verwandten mathematischen Methoden die Bedingungsvariationen der unabhängigen Variablen 'Methode' dar, deren jeweiliger Einfluß auf die abhängigen Variablen 'Klassifikationsleistung', 'inhaltlich-klinische Plausibilität des Modells' und 'praktische Anwendbarkeit der Klassifikationsregeln' untersucht werden soll. Aus dieser Überlegung ergibt sich eine unifaktorielle Versuchsplan-Anlage mit einer unabhängigen Variablen (UV) und drei abhängigen Variablen (AV).

3.1.2 Daten

Die den Berechnungen zugrunde liegenden Daten entstammen dem Köln-Bonner Früherkennungsprojekt (Klosterkötter, Hellmich & Schultze-Lutter, 2000; Klosterkötter, Hellmich, Steinmeyer & Schultze-Lutter, 2001; Schultze-Lutter, Klosterkötter & Steinmeyer, 2000), das

zum Ziel hatte, die Möglichkeit zur Vorhersage schizophrener Erstepisoden anhand von Basissymptomen prospektiv zu prüfen (vgl. Kap. 2.1.7.3).

3.1.2.1 Stichprobe

Die Stichprobe für das Köln-Bonner Früherkennungsprojekt wurde unter allen 695 Patienten ausgewählt, die vor 1991 in den Ambulanzen der Psychiatrischen Universitätskliniken Ulm-Weißenu, Lübeck, Bonn, Aachen und Köln mit der BSABS auf das Vorliegen von Prodromalsymptomen untersucht worden waren. Vor ihrer Übermittlung an eine dieser Ambulanzen hatten die meisten dieser Patienten schon mit Vorfeldeinrichtungen in Kontakt gestanden und waren zunächst von psychologischen Beratungsstellen oder Allgemeinärzten zu niedergelassenen Psychiatern, Neurologen oder Psychotherapeuten weitergeleitet worden. In deren Praxen hatten sie oft nacheinander verschiedene Diagnosen aus dem Bereich der Persönlichkeitsstörungen, Neurosen oder Erlebnisreaktionen zugeschrieben bekommen. Meist waren zum Zeitpunkt der Indexuntersuchung (t_1) mehrere Behandlungsversuche bereits mehr oder weniger erfolglos durchlaufen worden, und dieses unbefriedigende Ergebnis hatte zu dem Eindruck geführt, daß es sich um diagnostisch und therapeutisch besonders schwierige Fälle handelte. Die Überweisung an die Ambulanzen der genannten Kliniken war dementsprechend meist mit dem Ziel einer differenzierteren diagnostische Einschätzung sowie der Frage nach einer den Beschwerdebildern möglicherweise zugrundeliegenden beginnenden schizophrenen Störung erfolgt. Aus diesem Grunde wurden die Patienten in den Ambulanzen sehr ausführlich auf eine eventuell vorliegende (prä-)psychotische Symptomatik mit der BSABS und der PSE 9 untersucht.

Die Indexuntersuchung ergab bei einem Großteil der überwiesenen Patienten, daß in der Tat schon die diagnostischen Kriterien einer schizophrenen Störung zu sichern waren oder das jeweilige Beschwerdebild auf eine substanzinduzierte Störung, eine organisch bedingte Störung oder eine mentale Retardierung zurückzuführen war. Alle diese Fälle wurden von der Studie ausgeschlossen. Zudem blieben auch solche Patienten unberücksichtigt, die bei t_1 schon älter als 50 Jahre waren, da sie sich bei der geplanten Nachuntersuchung (t_2) bereits in einer stärker durch körperliche Erkrankungen oder Todesfälle belasteten Altersgruppe befunden hätten (s. Abb. 25).

Bei den verbliebenen 385 Fällen waren bei t_1 keine positiven oder negativen Schizophreniesymptome gefunden worden, weder im damaligen Querschnitt noch in der gesamten vorherigen Lebensgeschichte. Diese Patienten erhielten dem vorherrschenden klinischen Bild entsprechende Diagnosen (s. Abb. 26) und kehrten zumeist in die Betreuung ihrer überweisen-

den Ärzten und Psychotherapeuten zurück; nur ein kleiner Teil verblieb zur Behandlung in den entsprechenden Ambulanzen.

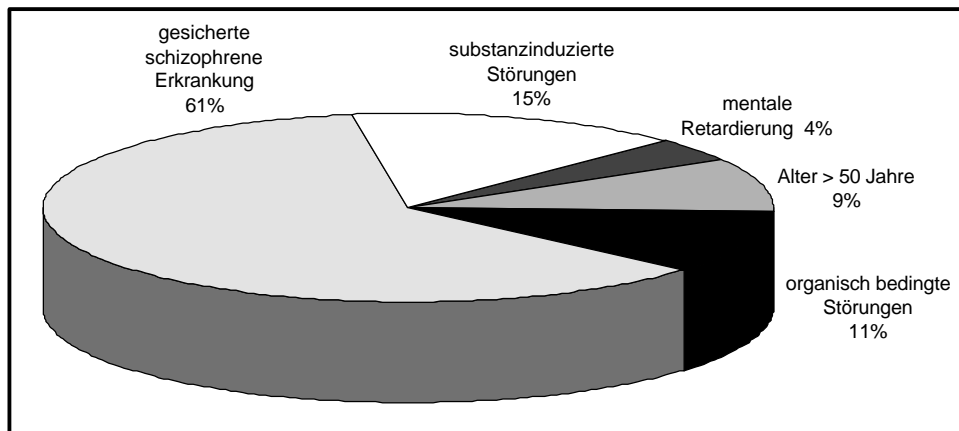


Abb. 25: Diagnoseverteilung und Anteil der über 50-jährigen der aus dem Projekt ausgeschlossenen Patienten (N = 310). Die Diagnosen sind entsprechend DSM-IV-Kriterien revidiert.

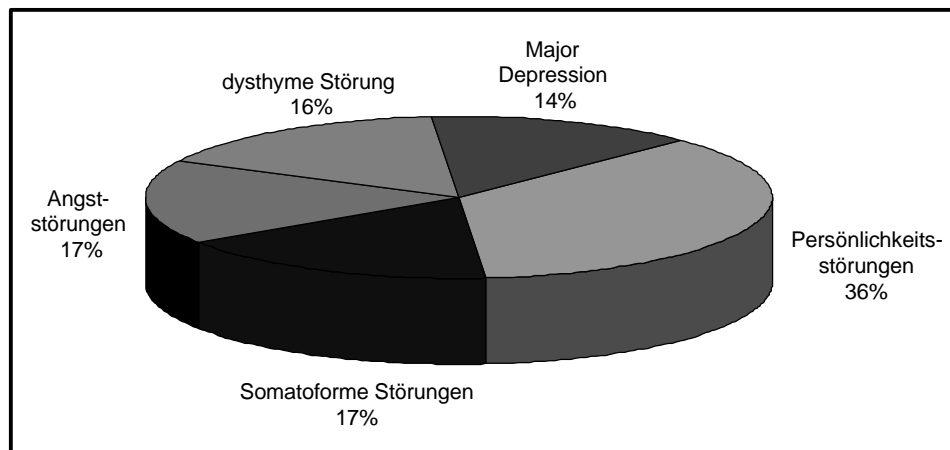


Abb. 26: Diagnoseverteilung der in das Projekt eingeschlossenen Patienten (N = 385). Die Diagnosen sind entsprechend DSM-IV-Kriterien revidiert.

Da aber ein Großteil dieser Patienten schon bei t_1 psychopathologische Veränderungen geboten hatten, die als Prodromalsymptome einer schizophrenen Störung verdächtig waren, wurde 1989 in Bonn mit ihrer Nachuntersuchung begonnen, die in Aachen fortgeführt und schließlich 1998 in Köln abgeschlossen wurde.

Das Ergebnis der Rekrutierungsbemühungen ist in Abbildung 27 dargestellt.

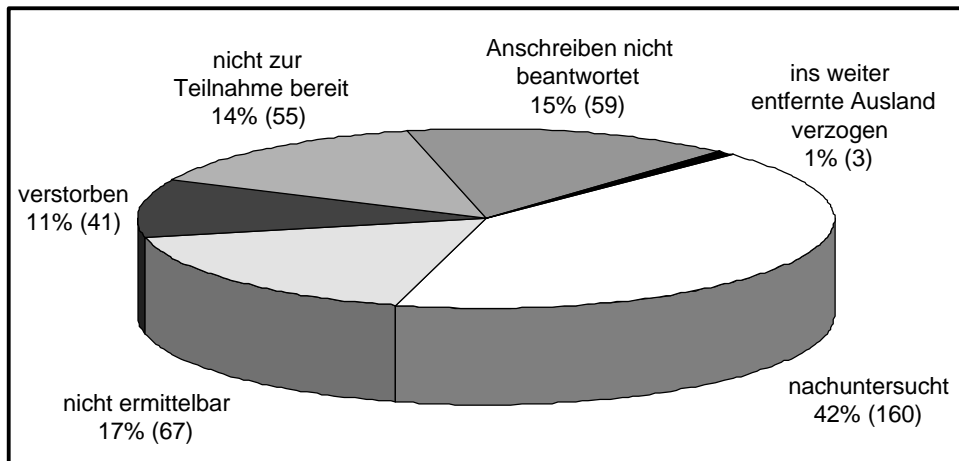


Abb. 27: Ergebnis der Rekrutierungsbemühungen um die in das Projekt eingeschlossenen Patienten (N = 385).

Insgesamt konnten so 160 Patienten nachuntersucht werden, die bei t_1 je nach dem vorherrschenden klinischen Bild Diagnosen vornehmlich aus dem Bereich der Neurosen und Persönlichkeitsstörungen erhalten hatten (s. Abb. 28). Hinsichtlich der vorläufigen auf dem aktuellen klinischen Bild zum Zeitpunkt der Indexuntersuchung beruhenden Diagnosen fanden sich keine signifikanten Unterschiede zwischen der nachuntersuchten und der nicht nachuntersuchten Gruppe, so daß hier nicht von einem systematischen Selektionsfehler zugunsten schwerer erkrankter Patienten ausgegangen werden kann (Klosterkötter et al., 2001). Auch hinsichtlich der als Prodromalsymptome betrachteten, bei der Indexuntersuchung geschilderten Beschwerden ergab sich kein signifikanter Unterschied zwischen der nachuntersuchten und der nicht nachuntersuchten Gruppe. Während in der nicht nachuntersuchten Gruppe 143 Patienten Prodromalsymptome nach BSABS-Kriterien genannt hatten und 82 nicht, waren es in der nachuntersuchten Gruppe 110 Patienten gewesen, die Prodromalsymptome gezeigt hatten, und 50, die keine berichteten. Auch hier ergibt sich somit kein Hinweis auf einen systematischen Selektionsfehler zugunsten schwerer erkrankter Fälle bei der Nachuntersuchung (Klosterkötter et al., 2001). Darüber hinaus fanden sich zwischen den beiden Gruppen auch keine signifikanten Unterschiede in der Alters- oder Geschlechtsverteilung. Damit kann die nachuntersuchte Stichprobe von 160 Patienten durchaus als repräsentativ für die gesamte Zielstichprobe von 385 Patienten angesehen werden (Klosterkötter et al., 2001).

Unabhängig von den bei t_1 vergebenen Diagnosen, waren von 110 der nachuntersuchten Patienten Störungen berichtet worden, die nach BSABS-Kriterien Prodromalsymptomen entsprachen. Für diese Fälle wurde die jeweilige Indexdiagnose deshalb nur als vorläufiger Kompromiß betrachtet und hypothetisch angenommen, daß sie sich bereits in der Prodromalphase einer schizophrenen Störung vor der psychotischen Erstmanifestation befanden.

Genau diese Hypothese wurde durch die Nachuntersuchung überprüft. Die übrigen 50 nachuntersuchten Fälle hatten bei t_1 keine auf Prodromalsymptome verdächtigten Beschwerden geboten, zeigten ansonsten aber hinsichtlich Diagnoseverteilung, Alter und Geschlecht bei t_1 keine signifikanten Unterschiede zu den 110 Patienten mit Prodromalsymptomen (Klosterkötter et al., 2001). Damit ließ der Vergleich der beiden Gruppen im Hinblick auf eine spätere Schizophrenieentwicklung die Beantwortung der Fragestellung zu, ob mit der BSABS als potentielle Prodromalsymptome identifizierte Störungen in der Tat nachfolgende psychotische Erstmanifestationen mit hinreichender Sicherheit vorhersagen können.

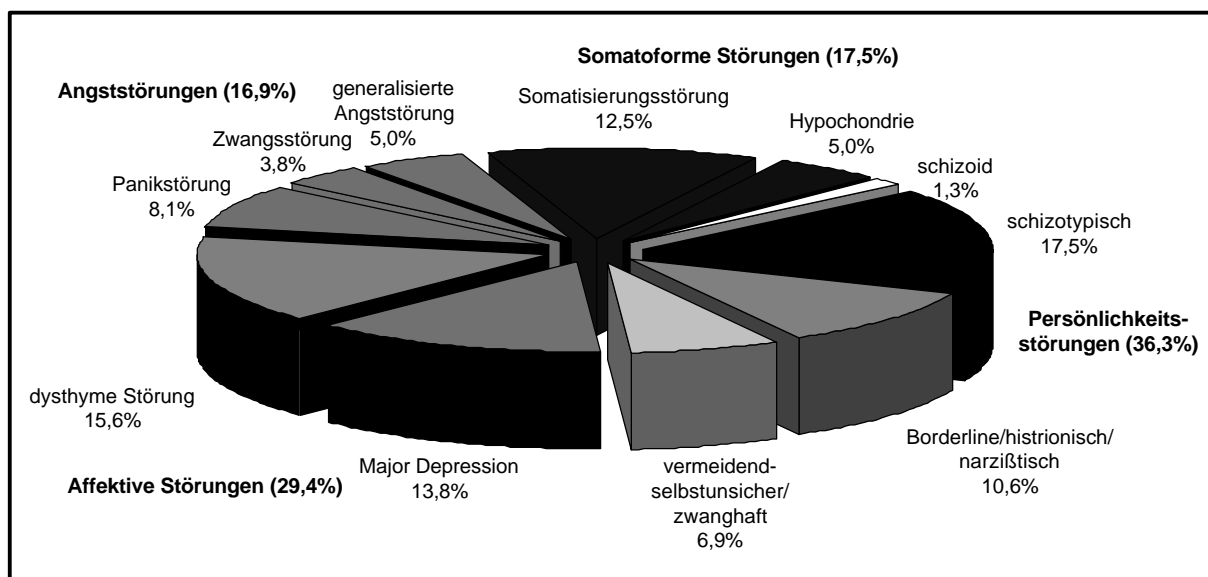


Abb. 28: Diagnoseverteilung der nachuntersuchten Patienten bei t_1 (N = 160). Die Diagnosen sind entsprechend DSM-IV-Kriterien revidiert.

Bei der im Durchschnitt etwa zehn Jahre nach der Indexuntersuchung erfolgten Nachuntersuchung zeigte sich, daß in der Tat 79 Personen eine schizophrene Erkrankung im Katamnesezeitraum entwickelt hatten; nur zwei von ihnen hatten bei der Indexuntersuchung keine Basissymptome geboten. Tabelle 12 zeigt die allgemeinen und anamnestischen Daten der nachuntersuchten Stichprobe.

3.1.2.2 Instrumente

Bei t_2 wurden dieselben psychometrischen Untersuchungsverfahren wie bei t_1 verwandt, BSABS und PSE 9, und das DSM-IV als Diagnosesystem zugrunde gelegt.

Tab. 12: Allgemeine und anamnestische Daten der nachuntersuchten Stichprobe (N = 160).

	Frauen (n = 76) Median Mittel ± s Range	Männer (n = 84) Median Mittel ± s Range	Gesamt (n = 160) Median Mittel ± s Range
Alter bei t₁ (in Jahren)	25,0 29,7 ± 11,4 15-50	26,0 28,9 ± 8,7 17-50	26,0 29,3 ± 10,0 15-50
Alter bei t₂ (in Jahren)	35,0 38,4 ± 12,4 21-72	37,5 40,6 ± 11,4 20-69	37,0 39,6 ± 11,9 20-72
Katamnesezeitraum (in Jahren)	5,5 8,6 ± 7,5 0-37	10,0 10,5 ± 7,6 0-30	7,8 9,6 ± 7,6 0-37
Prodromdauer (in Jahren, n = 79)	3,0 4,3 ± 3,6 0-13 (n = 36)	4,0 6,5 ± 5,8 1-22 (n = 43)	4,0 5,5 ± 5,1 0-22 (n = 79)

Die neunte Version der 'Present State Examination - PSE 9' (Wing, Cooper & Sartorius, 1982) ist ein halbstandardisiertes psychiatrisches Interview und dient der Beurteilung des psychischen Zustandes von erwachsenen Patienten mit Neurosen oder funktionellen Psychosen zum Untersuchungszeitpunkt sowie in den vorangegangenen vier Wochen. In der vorliegenden Untersuchung wurde der vorgesehene Erhebungszeitraum der PSE 9 von vier Wochen jedoch dahingehend ausgedehnt, daß das gesamte katamnestische Intervall zwischen t₁ und t₂ erfaßt wurde.

Neben der Anleitung zur Durchführung des diagnostischen Interviews, mit dem systematisch die wichtigsten diagnostischen Phänomene erhoben werden, sind Regeln zur Bewertung der einzelnen Symptome und zur Verdichtung dieser Informationen zu verschiedenen Syndromklassen im Manual enthalten. Die resultierenden deskriptiven Klassen sind mit den Kriterien der ICD-Diagnosen vergleichbar.

Das Gesamtverfahren besteht aus 140 Items, die in einem Glossar definiert sind. Obwohl dem Untersucher standardisierte, vorformulierte Fragen zur Verfügung stehen, steht es ihm frei, andere Formulierungen zu wählen, Zusatzfragen zu stellen oder die Reihenfolge der Fragen zu verändern, wenn dies in der aktuellen Untersuchungssituation angemessener erscheint. Jedes Symptom wird vom Untersucher auf sein Vorliegen hin überprüft, wobei sein klinisches Urteil Vorrang vor den Selbstangaben des Patienten hat. Die Bewertung der Symptome erfolgt zudem nach Ausprägung und Häufigkeit (0 = nicht vorhanden, 1 = vorhanden in mäßiger Ausprägung oder in schwerer Ausprägung, aber in weniger als der Hälfte der Zeit,

2 = vorhanden in schwerer Ausprägung in mehr als der Hälfte der Zeit, 8 = fraglich vorhanden, 9 = keine Beurteilung möglich). Die Symptome sind nach Gesichtspunkten der Zweckmäßigkeit in 20 sich teilweise überschneidende Sektoren gegliedert. Die ersten 17 Sektoren beziehen sich auf die Angaben des Patienten, die restlichen vier auf Verhalten, Affekt und Sprache während der Untersuchung.

Bezüglich der Schizophrenie deckt die PSE 9 alle diagnostisch relevanten Merkmale - insbesondere die Symptome ersten Ranges im Sinne Schneiders (1992) - sehr genau ab. Entsprechend setzt sich das schizophrene PSE 9-Kernsyndrom (Kategorieklasse S+) auch weitgehend aus diesen transkulturell stabilen (Jablensky et al., 1992) Schizophreniemerkmalen zusammen.

Die Reliabilität des PSE ist in verschiedenen Studien überprüft worden. In einer Übersicht kommen Domann und Klug (1979) zu dem Fazit einer insgesamt zufriedenstellenden Reliabilität des Verfahrens, obwohl die Vergleichbarkeit der Untersuchungen nicht generell gewährleistet sei, da verschiedene Versionen und Untersuchungsmethoden verwandt worden seien.

Die 'Bonner Skala für die Beurteilung von Basissymptomen - BSABS' (Gross, Huber, Klosterkötter & Linz, 1989) ist ebenfalls ein halbstandardisiertes Verfahren zur Erhebung und Dokumentation feiner, subklinischer psychopathologischer Beschwerden bei Patienten, bei denen die Diagnose einer Schizophrenie bereits gesichert ist oder der Verdacht auf ein Vorpostensyndrom oder Prodrom einer Schizophrenie besteht. Den ursprünglichen Ausführungen der Autoren folgend kann sie zudem bei idiopathischen Psychosen des 'schizoaffektiven Zwischenbereichs' (Janzarik, 1980) und affektiven Psychosen verwandt werden, da auch hier Basissymptome vorkommen können.

Obwohl die BSABS ein Fremdbeurteilungsverfahren ist, erfaßt sie als Basissymptome ausschließlich die vom Patienten selbst als Störung erlebten und berichteten Phänomene. Die BSABS hat ihre Ursprünge in der Heidelberger Checkliste von 1962, die anhand der Ergebnisse spätkatamnestischer Untersuchungen bei früheren Patienten der Heidelberger Psychiatrischen Universitätsklinik entwickelt wurde (Huber, 1966, 1968). Die 1987 publizierte Form der BSABS basiert auf einer Langzeitstudie an 202 schizophrenen Patienten (Huber, Gross & Schüttler, 1979), „die nach einer durchschnittlichen Verlaufsdauer von 22,4 Jahren seit mindestens fünf Jahren einigermaßen stabile reine Defizienzsyndrome zeigten“ (Huber, 1986, S. 132). Die Ergebnisse dieser Studie wurden als Bestätigung dafür gewertet, daß „Basissymptome ... vor der ersten Manifestation der Psychose in Prodromen und Vorpostensyndromen und nach der Remission der psychotischen Phasen ... die eigentlichen primären

Symptombildungen darstellen“ (Huber, 1986, S. 62). Die in der BSABS aufgeführten Basis-symptome, die nicht quantitativ, sondern nur qualitativ (0 = nicht vorhanden, 1 = fraglich vorhanden, 2 = vorhanden) erhoben werden, werden in fünf Hauptkategorien (A-E) mit insgesamt 98 Hauptitems (142 Einzelitems) eingeteilt:

- (A) Dynamische Defizienzen mit direkten Minussymptomen,
- (B) Dynamische Defizienzen mit indirekten Minussymptomen,
- (C) Kognitive Denk-, Wahrnehmungs- und Handlungsstörungen,
- (D) Coenästhesien
- (E) Zentral-vegetative Störungen inklusive Schlafstörungen und Intoleranz gegen bestimmte Substanzen.

In einer Zusatzkategorie (F) werden darüber hinaus sechs bewußte Bewältigungsversuche des Patienten gegenüber den selbstwahrgenommenen Basissymptomen festgehalten.

Das BSABS-Manual enthält für jedes Item neben den definitorischen Kriterien typische Aussagen von Patienten und Anregungen für gezielte Fragen des Untersuchers sowie Hinweise zur Abgrenzung von anderen Items. Da die BSABS nicht zur nosologischen Klassifizierung, sondern für psychopathologisch deskriptive Zwecke entwickelt wurde, enthält sie keine Normierungen für die einzelnen Kategorien oder den Gesamtwert.

Eine Untersuchung der Interrater-Reliabilität der BSABS über acht Untersucher, die zuvor alle an zehn Trainingssitzungen zur BSABS teilgenommen hatten, in 28 jeweils unterschiedlichen Paarkombinationen ergab eine mittlere Übereinstimmung der Untersucherpaare gemessen an Spearmans Rangkorrelation von $r_s = 0,808$ ($s = 0.102$), die als gut zu bewerten ist (Gross, Stassen, Huber & Klosterkötter, 1990).

Die in der vorliegenden Arbeit verwandte, gegenüber dem Originalinstrument auf 72 Items verkürzte Version der BSABS (vgl. Anhang C) beinhaltet Basissymptome der Kategorien A bis D, wobei auf die Differenzierung zwischen dynamischen Defizienzen mit direkter und indirekter Minussymptomatik (Kategorien A und B) verzichtet wurde und eher unspezifische Stufe-1-Basissymptome dieser beide Kategorien entfielen. Die Kategorien C und D wurden hingegen fast vollständig in diese Version aufgenommen.

BSABS und PSE 9 lassen sich aufgrund ihrer halbstrukturierten Form, die dem Untersucher eine Anpassung an die jeweilige Untersuchungssituation gestattet, gut miteinander kombinieren, denn beide Instrumente dienen der Fremdbeurteilung vornehmlich von Erlebnissymptomen und erfassen diese in phänomenologisch sehr differenzierter Form.

3.1.2.3 Untersuchungsdurchführung

Während die im regelmäßigen Kontakt mit den Ambulanzen verbliebenen Patienten leicht für eine Teilnahme an der Nachuntersuchung zu gewinnen waren, gelang der Einschluß derjenigen Patienten, die in die Betreuung durch niedergelassene Ärzte oder Psychotherapeuten zurückgekehrt waren oder die Therapie beendet hatten, oft erst nach wiederholten Versuchen der postalischen Kontaktaufnahme. Nach informierter Einwilligung führten jeweils zwei Mitarbeiter die Nachuntersuchung entweder im Rahmen eines Klinikbesuchs der ehemaligen Patienten oder eines Hausbesuchs in ihrer Heimatgemeinde durch. BSABS und PSE 9 kamen dabei immer gemeinsam zur Anwendung. Jeder der Untersucher hatte zuvor an einer Überprüfung der Urteilszuverlässigkeit über 10 Patienten teilgenommen und gemäß den Gütekriterien nach Schouten (1982) befriedigende Bias-korrigierte κ -Werte für die einzelnen BSABS und PSE-9-Subsyndrome mit einem Rang von 0.72 - 0.78 erzielt. Außerdem wurde streng darauf geachtet, daß die jeweiligen Rater bei der Nachuntersuchung blind für die Ergebnisse der Indexuntersuchung bei dem betreffenden Patienten waren. Die Zuweisung der Schizophreniediagnose erfolgte - ebenfalls ohne Kenntnis der Ergebnisse und Diagnosen bei t_1 - jeweils nach Abschluß der Untersuchungen zu t_2 durch zwei in der Handhabung des DSM-IV geübte Kliniker unter Berücksichtigung aller gewonnenen Befunde und Informationen zum Verlauf.

3.1.2.4 Eignung der Stichprobe

Der Datensatz erscheint aufgrund seines prospektiven Designs, der für diesen Forschungsbereich relativ umfangreichen Stichprobe und der fast gleichen Häufigkeit der beiden Ausprägungen der in die einzelnen Berechnungen als abhängigen Variable eingehenden Variable 'Ausgang' für die Fragestellung gut geeignet. Zudem hatten sich außer für die Indexdiagnose 'Persönlichkeitsstörung' keine signifikanten Korrelationen zwischen dem Ausgang und anderen potentiell konfundierenden Variablen gezeigt (s. Tab. 13).

Die signifikante Korrelation zwischen der Entwicklung einer schizophrenen Störung im katamnesticen Intervall und der Diagnose einer Persönlichkeitsstörung bei t_1 war, wie Häufigkeitsvergleiche der einzelnen Persönlichkeitsstörungen zeigten (Klosterkötter, Hellmich & Schultze-Lutter, 2000), auf ein signifikant häufigeres Vorkommen der schizotypen Persönlichkeitsstörung bei t_1 in dieser Gruppe zurückzuführen (χ^2 -Test, $df = 1$, $p < 0,001$). Da die schizotype Persönlichkeitsstörung dem schizophrenen Spektrum zugerechnet wird (Olin & Mednick, 1996; WHO, 1993), im ICD-10 als dazugehöriger Begriff auch die „prodromale Schizophrenie“ (WHO, 1993, S. 114) aufgeführt ist und die früheren DSM-III-R-

Prodromalsymptome einer Schizophrenie heute im DSM-IV weitgehend als Diagnosekriterien der schizotypen Persönlichkeitsstörung aufgeführt werden, ist dieser Zusammenhang nicht verwunderlich und zum Großteil sicherlich darauf zurückzuführen, daß das klinische Bild einer beginnenden Schizophrenie häufig am ehesten die diagnostischen Kriterien einer schizotypen Persönlichkeitsstörung erfüllen wird, insbesondere wenn eine sekundäre ängstliche und/oder depressive Symptomatik fehlen. Da somit die Diagnosestellung einer schizotypen Persönlichkeitsstörung eher als Folge der vorliegenden, potentiell prodromalen Symptomatik gewertet werden kann und damit nicht als eigenständige Einflußvariable auf die abhängige Variable 'Ausgang', wird sie nicht als weitere unabhängige Variable neben den einzelnen Basissymptomen oder als Störvariable im Sinne einer Kovariaten in die Berechnungen mit aufgenommen.

Tab. 13: Korrelationen zwischen der abhängigen Variablen 'Ausgang' und potentiell konfundierenden Variablen (Phi-, Spearman's- bzw. Pearsons Korrelationskoeffizient).

Variable	Korrelation
Allgemein anamnestische Daten	
Gestörte Heimverhältnisse (engl.: „broken home“)	-0,038
Höchster erreichter Ausbildungsstand	0,071
Familiäre Belastung mit schizophrenen Störungen	0,107
Familiäre Belastung mit bipolaren Störungen	0,101
Familiäre Belastung mit Angststörungen	0,065
Familiäre Belastung mit unklaren psychischen Störungen	-0,065
Geschlecht	-0,038
Alter bei t ₁	-0,079
Alter bei t ₂	0,021
Katamnesezeitraum	0,067
Indexdiagnosen, entsprechend DSM-IV revidiert	
Persönlichkeitsstörungen	0,191 *
affektive Störungen	-0,061
somatoforme Störungen	-0,093
Angststörungen	-0,078

* p < 0,05

3.1.2.5 Stichprobenteilung

Um eine Validierung der generierten Lösungen zu ermöglichen und Aufschluß über eine mögliche Überanpassung der Lösungen an die Berechnungsstichprobe zu erhalten, wurde

die Gesamtstichprobe der nachuntersuchten Patienten randomisiert in eine Berechnungsstichprobe von 120 Fällen und eine Validierungsstichprobe von 40 Fällen mit Hilfe der entsprechenden SPSS-Option zur Auswahl einer Zufallsstichprobe aufgeteilt. Diese Aufteilung wurde der Stichprobenhalbierung vorgezogen, um bei der gemessen an der Fallzahl großen Anzahl von in die einzelnen Berechnungen eingehenden unabhängigen Variablen (72 Basis-symptome) bei Verfahren wie etwa Regressions- und Diskriminanzanalysen, die nur eine begrenzte Menge unabhängiger Variablen zulassen, nicht schon bei der Berechnung eine zu große Vorselektion der eingehenden Variablen treffen zu müssen. Zudem hatten Studien zum Einfluß der Stichprobenauswahl und -größe auf Regressionsmodelle gezeigt, daß das Ausmaß der Übereinstimmung zwischen der Klassifizierungsleistung in der Berechnungs- und der Validierungsstichprobe mit steigendem Umfang der Berechnungsstichprobe zunimmt (Harrell et al., 1984, 1985), weshalb die Berechnungen an einer möglichst großen Stichprobe erfolgen sollten. Die allgemeinen anamnestischen Daten der beiden Stichproben sind in den Tabellen 14 und 15, die jeweilige Verteilung der Indexdiagnosen in den Abbildungen 29 und 30 wiedergegeben.

Tab. 14: Allgemeine und anamnestische Daten der Berechnungsstichprobe (n = 120).

	Frauen (n = 57) Median Mittel ± s Range	Männer (n = 63) Median Mittel ± s Range	Gesamt (n = 120) Median Mittel ± s Range
Alter bei t₁ (in Jahren)	26,0 30,5 ± 11,4 15-50	26,0 28,9 ± 9,2 17-50	26,0 29,7 ± 10,3 15-50
Alter bei t₂ (in Jahren)	34,0 38,0 ± 12,1 21-72	38,0 40,6 ± 11,9 20-69	36,0 39,3 ± 12,0 20-72
Katamnesezeitraum (in Jahren)	5,4 7,6 ± 5,8 0-24	9,8 10,2 ± 7,7 0-30	6,7 9,0 ± 6,9 0-30
Prodromdauer (in Jahren, n = 55)	3,5 4,6 ± 3,8 1-13 (n = 25)	4,0 7,3 ± 6,2 1-22 (n = 30)	4,0 6,1 ± 5,4 1-22 (n = 55)

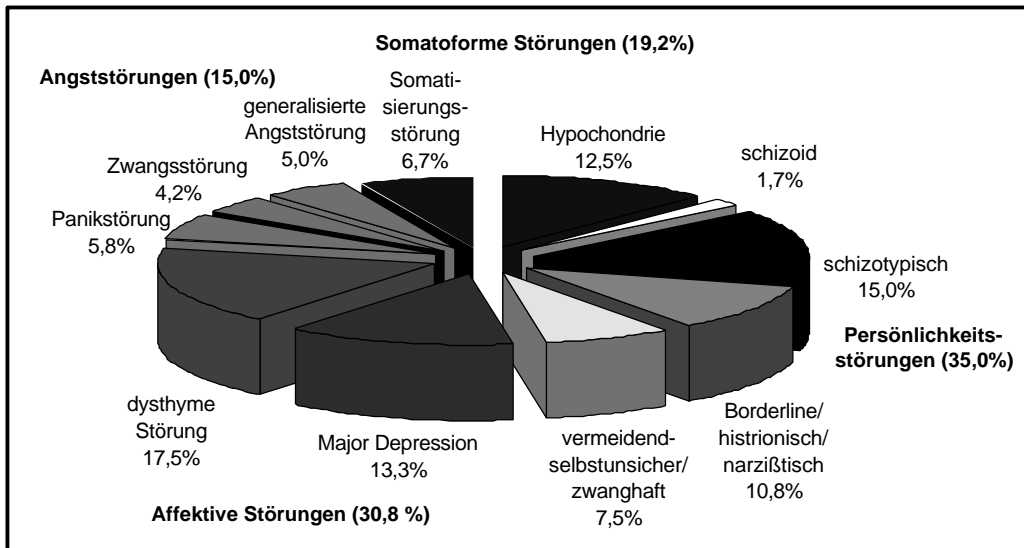


Abb. 29: Diagnoseverteilung der Berechnungsstichprobe bei t₁ (n = 120). Die Diagnosen sind entsprechend DSM-IV-Kriterien revidiert.

Tab. 15: Allgemeine und anamnestische Daten der Validierungsstichprobe (n = 40).

	Frauen (n = 19) Median Mittel ± s Range	Männer (n = 21) Median Mittel ± s Range	Gesamt (n = 40) Median Mittel ± s Range
Alter bei t₁ (in Jahren)	23,5 27,1 ± 11,1 16-50	27,0 29,1 ± 6,9 20-43	26,0 28,2 ± 9,1 16-50
Alter bei t₂ (in Jahren)	43,0 39,6 ± 13,4 21-63	37,0 40,6 ± 9,9 28-60	39,0 40,1 ± 11,6 21-63
Katamnesezeitraum (in Jahren)	6,0 11,3 ± 11,0 1-37	10,2 11,4 ± 7,2 1-29	9,9 11,3 ± 9,1 1-37
Prodromdauer (in Jahren, n = 24)	3,0 3,3 ± 2,9 0-8 (n = 11)	2,5 4,5 ± 4,1 1-13 (n = 13)	3,0 4,0 ± 3,6 0-13 (n = 24)

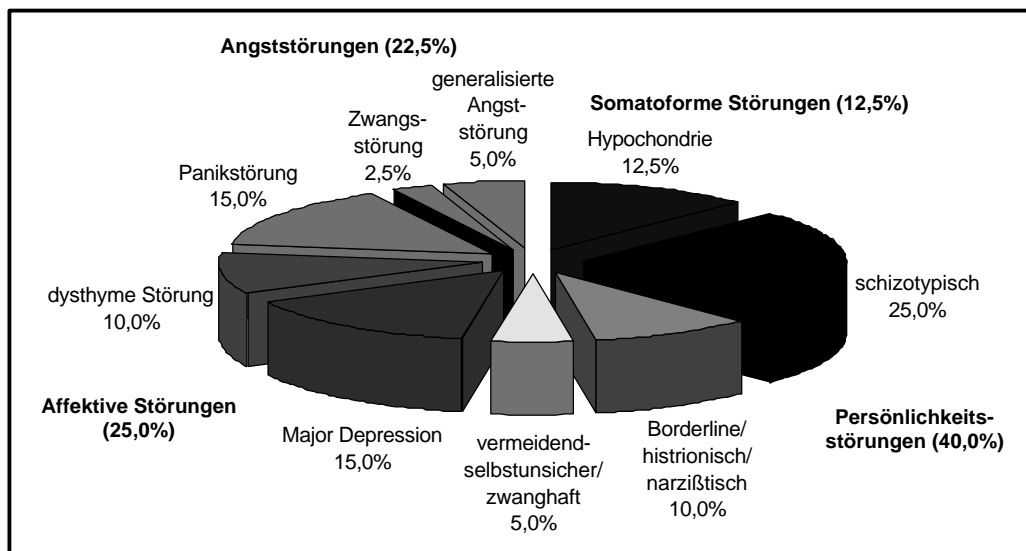


Abb. 30: Diagnoseverteilung der Validierungsstichprobe bei t_1 ($n = 40$). Die Diagnosen sind entsprechend DSM-IV-Kriterien revidiert.

3.2 VERSUCHSPLANUNG (2)

3.2.1 Bedingungsvariationen der unabhängigen Variable 'Methode'

Bei der kategorialen unabhängigen Variable 'Methode' (UV A) werden acht Bedingungsvariationen eingeführt³:

1. schrittweise Diskriminanzanalyse über die Merkmale, die auf Einzelitemebene einen signifikanten Gruppenunterschied in den Häufigkeiten aufweisen (2×2 - χ^2 -Test, $df = 1$, $p < 0,05$), wobei bei mehr als 26 Merkmalen die mit der höchsten Signifikanz ausgewählt werden, mit anschließender nonparametrischer Diskriminanzanalyse (DA-I),
2. schrittweise Diskriminanzanalyse über Merkmalscluster, die zuvor über eine hierarchische Clusteranalyse berechnet werden, mit anschließender nonparametrischer Diskriminanzanalyse (DA-C),
3. schrittweise logistische Regressionsanalyse über die Merkmale, die auf Einzelitemebene einen signifikanten Gruppenunterschied in den Häufigkeiten aufweisen (2×2 - χ^2 -Test, $df = 1$, $p < 0,05$), wobei bei mehr als 26 Merkmalen die mit der höchsten Signifikanz ausgewählt werden (LR-I),

³ Die Berechnungen erfolgen, soweit bei der Beschreibung der Methoden nicht anderweitig aufgeführt mit Hilfe der Standardversion des Programm-Pakets SPSS für Windows Version 7.5.

4. schrittweise logistische Regressionsanalyse über Merkmalscluster, die zuvor über eine hierarchische Clusteranalyse berechnet werden (LR-C),
5. Erkennung von Klassifikationsmustern auf der Grundlage des Grob-Mengen-Modells über alle Merkmale (GMM),
6. Search Partition Analysis über die gemessen an den positiven Produkt-Moment-Korrelationen 15 am höchsten mit dem Ausgang 'Übergang in Schizophrenie' korrelierten Merkmale (SPAN),
7. Ermittlung eines optimal trennenden Summenwertes für die den Selektionskriterien auf der Basis der diagnostischen Effizienzindizes (DEI) entsprechenden Merkmale,
8. Ermittlung eines optimal trennenden Summenwertes für die den Selektionskriterien auf der Basis diagnostischer Likelihood Ratios (DLR) entsprechenden Merkmale.

Die Auswahl der in die Diskriminanz- und Regressionsanalyse als Einzelvariablen eingehenden Merkmale erfolgt mit Hilfe einfacher Häufigkeitsvergleiche der Einzelmerkmale in den beiden Gruppen. Es werden diejenigen Merkmale ausgewählt, die in 2×2 - χ^2 -Tests bzw. bei einer erwarteten Zellenbesetzung $f_{(e)} \leq 5$ in Fishers Exaktem Test auf dem 5%-Niveau bei einem Freiheitsgrad signifikant werden. Trifft dies auf mehr als 26 Merkmale zu, so werden die 26 Merkmale ausgewählt, die die höchste Signifikanz aufweisen, da laut den Anwendungsempfehlungen von Backhaus und Mitarbeitern (1987) in beiden Verfahren die Zahl der Beobachtungen in der kleinsten Gruppe (hier die Gruppe mit dem Ausgang 'Übergang in Schizophrenie', $n = 55$) wenigstens doppelt so groß sein sollte wie die Anzahl der Variablen in der Analyse. Da dieser Schritt nicht zu statistischen Aussagen über Gruppenunterschiede führen soll, sondern nur der Vorselektion von Merkmalen gilt, also rein explorativ ist, wird hierbei auf eine Fehleradjustierung verzichtet.

Die Ermittlung der in die Diskriminanz- und Regressionsanalyse als Summenwerte eingehenden Cluster erfolgt mittels einer agglomerativen hierarchischen Clusteranalyse nach der Ward-Linkage-Methode. Das Ward-Verfahren ist ein eher konservatives Verfahren basierend auf Distanzmaßen, das weder zu dilatierenden, noch zu kontrahierenden Clusterbildungen neigt, das Heterogenitätsmaß mit zunehmender Fusionierung monoton ansteigen läßt und zur Bildung etwa gleich großer Cluster neigt. Untersuchungen haben gezeigt, daß das Ward-Verfahren im Vergleich zu anderen Algorithmen in den meisten Fällen sehr gute Partitionen findet und die Objekte den 'richtigen' Clustern zuordnet (Backhaus et al., 1987; Berg, 1981; Milligan, 1980, 1981). Für die Distanzmatrix werden die quadrierten Euklidischen Distanzen berechnet. Da ein hierarchisches Vorgehen keine Neuordnung einmal zugeordneter Variablen im Zuordnungsprozeß erlaubt, wird im Anschluß daran eine limitierte Clusteroptimierung entsprechend der höchsten Pearson-Korrelation zwischen Merkmal und Cluster sowie der inhaltlich-klinischen Plausibilität vorgenommen. Unter der Voraussetzung, daß die di-

chotome Variable wie hier 0/1-kodiert ist, entspricht der Korrelationskoeffizient nach Pearson dem punktbiserialen Korrelationskoeffizienten. Bei der Variablenreduktion über Struktur-aufdeckende Verfahren wird die Clusteranalyse gegenüber einer Hauptkomponentenanalyse bevorzugt, da der Vergleich von Marshall et al. (1994) gezeigt hat, daß eine schrittweise logistische Regressionsanalyse mit vorgeschalteter Clusteranalyse zu besseren Klassifikationsleistungen in der Validierungsstichprobe führte als die schrittweise logistische Regressionsanalyse mit vorgeschalteter Hauptkomponentenanalyse (vgl. Kap. 2.2.6).

Die schrittweise Diskriminanzanalyse erfolgt sowohl nach der Forward- als auch nach der Backward-Methode mit einem jeweiligen 5%-Signifikanzniveau für den Einschluß und Verbleib, um etwaige Unterschiede bei den Selektionsstrategien zu berücksichtigen. Treten hierbei Unterschiede auf, so wird die Selektion gewählt, die in der Berechnungsstichprobe zu einer insgesamt höheren Klassifizierungsleistung bzw. zu einer höheren Klassifizierungsleistung in der erkrankten Gruppe führt. Obwohl die schrittweise Diskriminanzanalyse eigentlich eine Normalverteilung der Daten voraussetzt, kann sie deskriptiv auch für nonparametrische Daten verwandt werden, wenn sich eine nonparametrische Analyse über die selektierten Merkmale anschließt (SAS Institute Inc., 1989). Die nonparametrische Diskriminanzanalyse der zuvor schrittweise selektierten Merkmale erfolgt mit der k-nearest-neighbour-Methode und verwendet die Mahalanobis Distanz als Proximitätsmaß. Die Zuordnung eines Objekts erfolgt in die nächstgelegene Gruppe. Die Diskriminanzanalysen werden mit dem Programm-Paket SAS Release 6.12, SAS/STAT, berechnet, da die SPSS 9.0-Standardversion keine Option zur Berechnung nonparametrischer Diskriminanzanalysen enthält.

Die schrittweise logistische Regressionsanalyse folgt der Wald-Methode mit sowohl 'Vorwärts'- als auch 'Rückwärts'-Selektion mit Zulassung einer Konstanten. Da die bereits in der Regressionsgleichung befindlichen Merkmale Einfluß auf den berechneten partiellen Korrelationskoeffizienten bzw. dessen Signifikanzniveau haben, werden beide Selektionsprozeduren berechnet, um mögliche Unterschiede in der Merkmalsselektion aufzudecken. Treten hierbei Unterschiede auf, so wird die Selektion gewählt, die in der Berechnungsstichprobe zu einer insgesamt höheren Klassifizierungsleistung bzw. zu einer höheren Klassifizierungsleistung in der erkrankten Gruppe führt. Als Einschluß- bzw. Ausschlußwert wird jeweils ein 5%-Signifikanzniveau definiert. Der Schwellenwert zur Klassifizierung in eine Gruppe wird auf $p = 0,5$ festgelegt.

Bei der Generierung von Vorhersageregeln auf Grundlage des Grob-Mengen-Ansatzes wird das Software-Paket 'DataLogic/R+1.5' (REDUCT System Inc., 1991-1994) verwandt. Es werden möglichst wenig und allgemein gültige, d.h. generelle, deterministische Regeln angestrebt, so daß eine hohe 'roughness' (0,90) und eine geringe 'rule precision' (0,10) voreingestellt wird (vgl. Kap. 2.2.4.2).

Als eine weitere alternative Methode zur Generierung von Vorhersageregeln, die der Booleschen Logik folgen, wird das Programm 'Search Partition ANalysis - SPAN' (Marshall, 1997) eingesetzt (vgl. Kap. 2.2.3.1). Die Messung der Partitionsreinheit erfolgt über die Entropie. Als Apriori-Wahrscheinlichkeit wird die Gruppenverteilung der Berechnungsstichprobe zugrunde gelegt (Option 'data priors'). Um die iterative Suche auf ein mit einem Standard-PC verarbeitbares Maß zu beschränken, wird die in einem Suchschritt der iterativen Suche erlaubte Attributanzahl pro Partition sowie die Anzahl der Partitionen auf jeweils 2 beschränkt, wobei die in einem Suchschritt gefundene Partition als neues Attribut in den jeweils nächsten eingehen kann ($q=2$, $p_1=2$, $p_2=1$). Die Straffunktion β wird gemäß den Anwendungsempfehlungen initial gleich 0 gesetzt (Marshall, 1995, p. 2653), für den 'balancing parameter' γ wird 0,5 gewählt, d.h. eine möglichst gleichmäßige Verteilung der Fälle in beide Gruppen der abhängigen Variable angestrebt, die dem zu erwartenden Ergebnis bei Zufallsklassifizierung Rechnung trägt.

Bei der Auswahl von Merkmalen nach den 'klassischen' diagnostischen Effizienzindizes wird sowohl Sensitivität, positive prädiktive Stärke als auch der Prozentsatz falsch-positiver Vorhersagen berücksichtigt. In Abwesenheit verbindlicher Kriterien, werden hierbei etwas liberalere Werte als die für die Diagnose einer akuten Schizophrenie diagnostisch relevanten Symptome geltenden angelegt: Sensitivität ≥ 0.25 , positive prädiktive Stärke ≥ 0.65 und Prozentsatz falsch-positiver Vorhersagen $\leq 10\%$. Als cut-off wird derjenige Summenwert gewählt, der zu den meisten korrekten Klassifizierungen in der Berechnungsstichprobe führt.

Bei der Auswahl von Merkmalen nach den diagnostischen Likelihood Ratios wird in Abwesenheit verbindlicher Selektionskriterien eine Steigerung des Erkrankungsrisikos bei vorhandenem Merkmal um den Faktor vier gefordert, d.h. Merkmale mit einer positiven diagnostischen Likelihood Ratio ≥ 4 ausgewählt. Als cut-off wird derjenige Summenwert gewählt, der zu den meisten korrekten Klassifizierungen in der Berechnungsstichprobe führt.

3.2.2 Versuchsplan-Anlage (2)

Aus den Operationalisierungen der unabhängigen Variable 'Methode' ergibt sich eine unifaktorielle Versuchsplan-Anlage mit acht kategorialen Ausprägungen der unabhängigen Variablen und drei abhängigen Variablen (s. Tab. 16).

Tab. 16: Versuchsplan-Anlage VPL-A1 (Hager, 1987).

Unabhängige Variable 'Methoden' (UV A)	AV1	AV2	AV3
A1: schrittweise und nonparametrische Diskriminanzanalyse über signifikante Merkmale (DA-I)	Prozentsatz richtiger Klassifizierungen in der Validierungsstichprobe	Plausibilität der Vorhersageregeln(n)	Anwendbarkeit der Vorhersageregeln(n)
A2: schrittweise und nonparametrische Diskriminanzanalyse über Cluster (DA-C)			
A3: schrittweise logistische Regression über signifikante Merkmale (LR-I)			
A4: schrittweise logistische Regression über Cluster (LR-C)			
A5: Grob-Mengen-theoretisch basierte Analyse (GMM)			
A6: Search Partition Analysis (SPAN)			
A7: Selektion nach diagnostischen Effizienzindizes und Einführung eines cut-offs (DEI)			
A8: Selektion nach diagnostischen Likelihood Ratios und Einführung eines cut-offs (DLR)			

Die Güte der Vorhersagegenauigkeit der einzelnen generierten Modelle in der Validierungsstichprobe wird in Anlehnung an Harrell und Mitarbeiter (1982) über die Reliabilität bestimmt, d.h. die Übereinstimmung zwischen vorhergesagtem und beobachtetem Ausgang. Die hierbei erhobene intervallskalierte abhängige Variable 'Klassifikationsleistung' (AV1) berechnet sich als der Prozentsatz korrekter Gruppenzuordnungen mit Hilfe des mit der jeweiligen Methode generierten Klassifizierungsalgorithmus' in der Validierungsstichprobe. Auf eine Prüfung der Diskriminationsfähigkeit über die Fläche unter der Receiver Operating Characteristic (ROC) Kurve muß hier verzichtet werden, da hierfür zumindest ein Ordinalskalenniveau der zu überprüfenden Zuordnungen vorliegen muß, die von SPAN und DataLogic generierten Klassifizierungsregeln jedoch nur im Hinblick auf ihre korrekte Zuordnung, also dichotom interpretierbar sind.

Die ordinalskalierte abhängige Variable 'inhaltlich-klinische Plausibilität des Modells' (AV2) wird über Expertenratings erhoben. Hierzu werden die generierten Vorhersageregeln in ausformulierter Form dargeboten und die Lösungen von den sieben ärztlichen und psychologischen wissenschaftlichen Mitarbeitern des Früherkennungs- und -therapie-Zentrums für psychotische Krisen, FETZ, der Kölner psychiatrischen Universitätsklinik aufgrund ihrer klinischen und wissenschaftlichen Erfahrung hinsichtlich ihrer inhaltlichen Plausibilität in eine Rangordnung gebracht (vgl. Anhang D). Entsprechend wird auch bei der Beurteilung der dritten ebenfalls ordinalskalierten abhängigen Variablen, 'Anwendbarkeit der Klassifizierungsregeln' (AV3), vorgegangen (vgl. Anhang D).

3.2.3 Ableitung von empirischen Vorhersagen

Obwohl in den letzten Jahren eine Vielzahl von Verfahren zur Generierung von Vorhersagemodellen mit einem reduzierten nominalskalierten Variablensatz entwickelt worden sind (vgl. Kap. 2.2.1 bis 2.2.5), wurden Untersuchungen zum Vergleich der Klassifizierungsleistungen dieser Verfahren bislang eher vernachlässigt oder sind auf Vergleiche von neu entwickelten Verfahren, etwa Partitionierungs- oder Grob-Mengen-Modelle, mit bereits etablierten Verfahren, wie Regressions- oder Diskriminanzanalysen, begrenzt (vgl. Kap. 2.2.6). Die Ergebnisse bisheriger Methodenvergleiche sind zudem zum Teil widersprüchlich und lassen sich wegen der unterschiedlichen zugrunde liegenden Stichproben nur schwer miteinander vergleichen. Dennoch zeichnen sich einige generelle Trends ab: So ist zu erwarten, daß die logistische Regressionsanalyse zu einer besseren Klassifikationsleistung führt als die Diskriminanzanalyse, wobei eine vorgeschaltete Clusteranalyse zur Reduktion der in die Analyse eingehenden Variablen noch einmal den Prozentsatz korrekter Klassifizierungen im Vergleich zur Analyse über Einzelmerkmale erhöhen sollte. Darüber hinaus weisen eine Reihe bisheriger Ergebnisse darauf hin, daß Partitionierungsverfahren zumindest zu einer der schrittweisen logistischen Regressionsanalyse vergleichbaren Klassifizierungsleistung führen und daß Grob-Mengen-basierte Verfahren bei Einbeziehung der Grenzregion der Diskriminanzanalyse überlegen sind (vgl. Kap. 2.2.6). Bei der Selektion von Merkmalen nach ihrer diagnostischen Effizienz ist eine Überlegenheit der diagnostischen Likelihood Ratios über die 'klassischen' diagnostischen Effizienzindizes zu erwarten, da erstere einen direkteren Bezug zum Prädiktionswert eines Tests bzw. Symptoms besitzen (Leisenring & Pepe, 1998).

Neben der objektiv meßbaren Klassifizierungsleistung der verschiedenen Verfahren finden immer wieder theoretische Überlegungen in die Diskussion um die Verfahren Eingang, die auf die Plausibilität und Handhabbarkeit der generierten Lösungen abzielen. Hierbei wurde vielfach betont, daß es als Vorteil anzusehen sei, wenn der Klassifizierungsalgorithmus der prognostischen Entscheidungsbildung im klinischen Alltag nahe stehe und somit leicht verständlich sei (Cook & Goldman, 1984; Koss & Feinstein, 1971; Leisenring & Pepe, 1998; Marshall, 1986). Die Diskussion um die Plausibilität und Praktikabilität der aus den unterschiedlichen Verfahren resultierenden Klassifizierungsregeln ist bisher jedoch eine rein theoretische, eine empirische Untersuchung der Einschätzung generierter Regeln hinsichtlich dieser Eigenschaften von erfahrenen Klinikern ist nicht vorgenommen worden.

Treffen die in der Diskussion um die klinische Plausibilität der generierten Klassifikationsregeln gerade für die alternativen Verfahren angeführten Argumente zu, so sollten Selektions- und Partitionierungsverfahren als plausibler als Regressions- und Diskriminanzverfahren beurteilt werden und im Falle der Search Partition Analysis auch dem Grob-Mengen-Modell überlegen sein, da in letzteres auch die Abwesenheit eines Symptoms innerhalb einer Ent-

scheidungsregel für die Vorhersage der Erkrankung eingehen kann. Die Selektionsmodelle über die diagnostischen Gütekriterien sollten sich hinsichtlich der Beurteilung ihrer klinischen Plausibilität weitgehend entsprechen.

Da im Hinblick auf die Anwendbarkeit zu erwarten ist, daß einfache und übersichtliche Verfahren bevorzugt werden, wird vorausgesagt, daß die Einführung eines cut-offs, wie dies bei der Selektion nach diagnostischen Gütekriterien der Fall ist, allen anderen Modellen vorgezogen wird, wobei die Lösung mit der geringeren Anzahl von Variablen noch einmal positiver beurteilt wird. Entsprechend wird auch erwartet, daß die einfacheren Regeln der Search Partition Analyse gegenüber den komplexeren Regeln der Grob-Mengen-Modelle als anwenderfreundlicher eingestuft werden. Zudem wird angenommen, daß die loglinearen Regressionsgleichungen den eher dimensional ausgelegten Lösungsalgorithmen der nonparametrischen Diskriminanzanalysen hinsichtlich ihrer Handhabbarkeit vorgezogen werden, wobei auch hier die jeweilige Lösung mit der geringeren Anzahl von eingehenden Variablen, die voraussichtlich die Berechnungen über die Cluster sein wird, präferiert werden.

3.2.4 Ableitung von statistischen Vorhersagen

Aus den empirischen Vorhersagen ergeben sich folgende gerichtete statistische Vorhersagen für die AV1, 'Klassifizierungsleistung' (KL), die sich über die erwarteten Häufigkeiten richtiger Gruppenzuordnungen in der Validierungsstichprobe, $e f_{[KL(k)]}$ mit k = Ausprägung der UV, ausdrücken lassen:

1. $e f_{[KL(SPAN)]} \geq e f_{[KL(LR-C)]} > e f_{[KL(LR-I)]} > e f_{[KL(DA-C)]} \geq e f_{[KL(DA-I)]}$ und
2. $e f_{[KL(GMM)]} > e f_{[KL(DA-I)]}$ und
3. $e f_{[KL(DLR)]} \geq e f_{[KL(DEI)]}$.

Für die zweite AV, 'Plausibilität der Vorhersageregeln' (P), resultieren entsprechend den empirischen Vorhersagen (vgl. Kap. 3.2.3) die nachfolgenden statistischen Vorhersagen für die erwarteten mittleren Ränge, $e \bar{R}_{P(k)}$ mit k = Ausprägung der UV, wobei ein niedriger Rang eine hohe Plausibilität ausdrückt:

1. $e \bar{R}_{P(SPAN)} < e \bar{R}_{P(GMM)} < e \bar{R}_{P(LR-C)} \leq e \bar{R}_{P(LR-I)} \leq e \bar{R}_{P(DA-C)} \leq e \bar{R}_{P(DA-I)}$ und
2. $e \bar{R}_{P(DLR)} \cong e \bar{R}_{P(DEI)} \leq e \bar{R}_{P(LR-C)} \leq e \bar{R}_{P(LR-I)} \leq e \bar{R}_{P(DA-C)} \leq e \bar{R}_{P(DA-I)}$.

und für die AV3, 'Anwendbarkeit der Vorhersageregeln' (A) werden folgende statistische Vorhersagen für die erwarteten mittleren Ränge, $e \bar{R}_{A(k)}$ mit k = Ausprägung der UV, abgeleitet, wobei ein niedriger Rang eine Anwendbarkeit ausdrückt:

1. $e \bar{R}_{A(\text{DLR})} \cong e \bar{R}_{A(\text{DEI})} < e \bar{R}_{A(\text{LR-C})} \leq e \bar{R}_{A(\text{LR-I})} < e \bar{R}_{A(\text{DA-C})} \leq e \bar{R}_{A(\text{DA-I})}$ und
2. $e \bar{R}_{A(\text{DLR})} \cong e \bar{R}_{A(\text{DEI})} < e \bar{R}_{A(\text{SPAN})} < e \bar{R}_{A(\text{GMM})}$.

3.2.5 Ableitung von statistischen Hypothesen

Die statistischen Vorhersagen lassen sich für die drei abhängigen Variablen in die folgenden inferenzstatistisch direkt prüfbaren statistischen Hypothesen überführen:

Für die AV1 'Klassifizierungsleistung' resultieren

1. $H_1: p(\chi^2_{(7;95\%)}) \leq 0,05$ für den Vergleich aller acht Methoden und
2. $H_1: p_m(\chi^2_{(1;95\%)}) \leq 0,05$ für mindestens einen der (8 über 2) Paarvergleiche m.

Für die AV2 'Plausibilität der Vorhersageregeln(n)' und die AV3 'Anwendbarkeit der Vorhersageregeln(n)', die jeweils nur von sieben Ratern eingeschätzt werden, sind die Daten aufgrund der geringen Stichprobenzahl nur deskriptiv, aber nicht inferenzstatistisch auswertbar, so daß hier auch auf die Auflistung der multiplen, genauer der sich für die AV2 und AV3 jeweils ergebenden (8 über 2) = 28 Paarvergleiche der Mediane und Modalwerte der einzelnen Verfahren verzichtet wird.

3.2.6 Testplanung

Zur statistischen Prüfung der ersten H_1 , der Annahme eines Unterschiedes zwischen den Häufigkeiten der einzelnen Bedingungsvariationen 'Klassifizierungsleistung der Methoden in der Validierungsstichprobe', wird ein 1-dimensionaler χ^2 -Test mit $k - 1 = 7$ Freiheitsgraden gerechnet und bei zweiseitiger Fragestellung auf dem 5%- α -Niveau getestet ($\chi^2_{\text{krit}(7; 95\%)} = 14,0671$). Hiermit läßt sich bei einer Teststärke = 0,90 ein nach den Konventionen von Cohen (1977) etwa mittlerer Effekt von $\delta = 0,5110$ aufdecken.

Die 28 Paarvergleiche der AV1 'Klassifizierungsleistung' über die acht Bedingungsvariationen der UV 'Methode' werden über 2×2 - χ^2 -Tests mit jeweils einem Freiheitsgrad und einem $\alpha = 5\%$ bei einseitiger Fragestellung kalkuliert ($\chi^2_{\text{krit}(1; 90\%)} = 2,7055$). Der 2×2 - χ^2 -Test entspricht der Überprüfung der Differenz zweier Prozentwerte aus unabhängigen Stichproben. Obwohl die Erhebungen der AV1 an ein und derselben Stichprobe erfolgen, werden die Bedingungsvariationen der UV als voneinander unabhängig betrachtet, so daß der 2×2 - χ^2 -Test zur Anwendung gelangen kann.

Die in der Literatur verfügbaren Methodenvergleiche (vgl. Kap. 2.2.6) lassen erwarten, daß sich die Klassifizierungsleistungen der einzelnen Methoden nicht sehr deutlich bzw. signifikant unterscheiden, d. h. der aufzudeckende Effekt δ gemäß den Konventionen von Cohen (1977) eher klein ($\delta < 0,2000$) ist. Die für die Validierungsstichprobe verfügbare Stichprobengröße erlaubt mit $N = 40$ bei einem $\alpha = 5\%$ bei einseitiger Testung und einer Teststärke von 0,90 jedoch nur die Aufdeckung eines sehr großen Effekts von $\delta = 0,9252$ durch die einzelnen 2×2 - χ^2 -Tests. Daher können statistisch signifikante Unterschiede in den Paarvergleichen kaum erwartet werden, und es wird auf die bei der Verwendung des schwachen Kriteriums strenggenommen notwendige Adjustierung des α -Fehlers (Hager, 1987) bewußt verzichtet, da sich dadurch der bei der vorgegebenen Stichprobengröße maximal aufzudeckende Effekt auf $\delta > 1,00$ erhöhen würde.

4 Ergebnisse

4.1 Ergebnisse der vorgeschalteten Analysen

4.1.1 Häufigkeitsvergleich

Häufigkeitsvergleiche der einzelnen unabhängigen Variablen in den beiden Outcome-Gruppen der abhängigen Variablen zeigten einen signifikanten Gruppenunterschied für 20 der 72 BSABS-Items auf dem 5%-Signifikanzniveau (s. Tab. 17; Anhang E enthält die Ergebnisse der Häufigkeitsvergleiche für alle 72 unabhängigen Variablen).

Diese 20 Einzelitems gehen entsprechend der Versuchsplanung als unabhängige Variablen in die schrittweise Diskriminanzanalyse (vgl. Kap. 4.2.1) bzw. die schrittweise logistische Regressionsanalyse (vgl. Kap. 4.2.3) ein.

4.1.2 Clusteranalyse

Als zweite, alternative Vorgehensweise zur Reduktion der in die schrittweise Diskriminanzanalyse bzw. die schrittweise logistische Regressionsanalyse eingehenden unabhängigen Variablen wurde eine hierarchische Clusteranalyse nach der Ward-Linkage-Methode über die Einzelvariablen gerechnet (s. Anhang F).

Tab. 17: Ergebnis der Häufigkeitsvergleiche der Einzelvariablen in den beiden Outcome-Gruppen der Berechnungsstichprobe (n=120): Signifikante Variablen

BSABS-Item-Nr.	Basissymptom	Häufigkeit in G ₀	Häufigkeit in G ₁	χ^2 -Wert ^a	p ^b
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	4	14	8,704	0,003
C.1.1	Gedankeninterferenz	6	24	18,808	0,000
C.1.2	Zwangähnliches Perseverieren zurückliegender Vorgänge	8	19	8,449	0,004
C.1.3	Gedankendrängen, -jagen	3	21	20,979	0,000
C.1.4	Blockierung des jeweiligen Gedankengangs	9	20	8,242	0,004
C.1.6	Störung der rezeptiven Sprache	4	22	20,108	0,000
C.1.7	Störung der expressiven Sprache	5	13	5,940	0,015
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	3	13	9,328	0,002
C.1.17	'Subjekt-Zentrismus', Eigenbeziehungstendenz	6	21	14,320	0,000
C.2.2 s1	Überempfindlichkeit gegen Licht oder visuelle Reize	2	12	10,154	0,001
C.2.2 s2	Photopsien	2	9	6,316	0,012
C.2.3 s5	Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	0	7	8,785	0,003
C.2.3 s6	Wahrnehmungsveränderungen am eigenen Gesicht	3	12	8,061	0,005
C.2.5 s1	Veränderung der Intensität / Qualität von Gehörs wahrnehmungen	5	11	3,905	0,048
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	4	13	7,488	0,006
C.2.8	Sensorische Überwachheit	1	11	11,282	0,001
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	0	7	8,785	0,003
C.2.11	Derealisation	7	14	4,450	0,035
C.3.1	Motorische Interferenz	0	6	7,464	0,008
C.3.3	Automatismenverlust	3	12	8,061	0,005

G₀: Gruppe der Patienten ohne Entwicklung einer schizophrenen Störung im Katamnese-Intervall

G₁: Gruppe der Patienten mit Entwicklung einer schizophrenen Störung im Katamnese-Intervall

^a 2x2- χ^2 -Test, df = 1

^b Asymptotische bzw. exakte Signifikanz bei Zellenbesetzungen mit einer erwarteten Häufigkeit $f_{(e)} < 5$ bei 2-seitiger Testung

Ausgehend von dem Elbow-Kriterium beim Ward-Verfahren, bei dem die Fehlerquadratsumme gegen die Clusterzahl in einem Koordinatensystem abgetragen und die Entscheidung für eine Clusteranzahl getroffen wird, die noch vor den Fusionen mit relativ starkem Heterogenitätszuwachs liegt (Backhaus et al., 1987), wurde eine auch inhaltlich weitgehend plausible Lösung von sieben Clustern ausgewählt (s. Anhang G). Da die hierarchische Clusteranalyse keine Umgruppierung einmal zugeordneter Variablen zulässt, wurden auf der Basis von Pearsons Korrelationskoeffizienten zwischen den Variablen und den generierten Clustern sowie inhaltlich-klinischer Überlegungen begrenzte Neuordnungen von insgesamt elf Variablen in zwei Schritten vorgenommen. Diese führten zu der in Tabelle 18 dargestellten Variablenzuordnung zu den sieben Clustern. Mit Ausnahme von zwei Variablen, 'Scheinbewegungen von Wahrnehmungsobjekten' (C.2.3 s7) und 'Doppelt-, Schief-, Schräg- und Verkehrtsehen' (C.2.3 s8), bei denen inhaltliche Gründe für ihre Zuordnung entscheidend waren, korrelierten bei dieser Lösung alle Variablen nicht nur hoch signifikant, sondern auch am höchsten mit ihrem eigenen Cluster (s. Anhang H). Für die Variablen 'Auflösung der Geradlinigkeit gegenständlicher Konturen' (C.2.3 s10) und 'kinästhetische Sensationen' (D.10), die in der Berechnungsstichprobe in keiner der beiden Gruppen auftraten (s. Anhang E), erfolgte die Zuordnung allein aufgrund inhaltlich-klinischer Kriterien. Diese sieben Cluster gehen mit ihrem jeweiligen Summenwert in die schrittweisen Diskriminanz- und Regressionsanalysen ein (vgl. Kap. 4.2.2 und 4.2.4).

Tab. 18: Variablenzuordnung zu den Clustern nach der zweiten Umgruppierung.

BSABS-Item-Nr.	Basissymptom
Überbeanspruchung (Cluster BS1)	
A.5	Mangelnde Entscheidungsfähigkeit, Entschlußschwäche, Unschlüssigkeit
A.6.4	Minderung des Kontaktbedürfnisses
A.7.1	Minderung der Kontaktfähigkeit bei vorhandenem Kontaktwunsch
A.7.2	Störung des In-Erscheinung-Tretens
A.8.1 + B.1.2	Minderung der psychischen Belastungsfähigkeit gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen
A.8.2 + B.1.3	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen
A.8.3 + B.1.4	Minderung der psychischen Belastungsfähigkeit gegenüber Arbeit unter Zeitdruck oder gegenüber rasch wechselnden unterschiedlichen Anforderungen
B.1.1	Minderung der psychischen Belastungsfähigkeit gegenüber körperlicher und/oder psychischer arbeitsmäßiger Beanspruchung
B.2.1	Erhöhte Beeindruckbarkeit durch alltägliche Ereignisse
B.2.2	Erhöhte Beeindruckbarkeit durch Verhaltensweisen anderer, die den Patienten persönlich betreffen
B.2.3	Erhöhte Beeindruckbarkeit durch „fremdes Leid“

Dynamische Defizite (Cluster BS2)

- A.6.1 Veränderung von Grundstimmung und emotionaler Resonanzfähigkeit
- A.6.3 Abschwächung bejahender Fremdwert- und Sympathiegefühle
- C.1.5 Störung der Konzentrationsfähigkeit

Wahrnehmungs- und Handlungsstörungen (Cluster BS3)

- C.1.15 Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen
- C.2.1 s1 Verschwommen- und Trübsehen
- C.2.1 s3 Partielles Sehen
- C.2.2 s1 Überempfindlichkeit gegenüber Licht / bestimmten visuellen Wahrnehmungsobjekten
- C.2.2 s2 Photopsien
- C.2.3 s1 Porropsie, Nahsehen
- C.2.3 s2 Mikro- und Makropsie
- C.2.3 s3 Metamorphopsie
- C.2.3 s4 Veränderungen des Farbensehens, Farbigsehen
- C.2.3 s5 Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer
- C.2.3 s7 Scheinbewegungen von Wahrnehmungsobjekten
- C.2.3 s8 Doppelt-, Schief-, Schräg- und Verkehrt-Sehen
- C.2.3 s9 Störung der Schätzung von Entfernungen / Größe von Gegenständen
- C.2.3 s10 Auflösung der Geradlinigkeit gegenständlicher Konturen
- C.2.4 s1 Überempfindlichkeit gegenüber Geräuschen, Lärm / allgemein akustischen Reizen
- C.2.4 s2 Akoasmen
- C.2.5 s1 Veränderungen von Intensität und/oder Qualität von Gehörs wahrnehmungen
- C.2.6 Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem, sensiblem Gebiet
- C.2.8 Sensorische Überwachheit
- C.2.10 Störung der Kontinuität der Wahrnehmung der eigenen Handlungen
- C.3.1 Motorische Interferenz, Automatosesyndrom
- C.3.2 Motorische Blockierung, Bannungszustände
- C.3.3 Verlust automatisierter Fertigkeiten (Automatismenverlust)

Entfremdung (Cluster BS4)

- A.6.2 Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten
- B.3.1 Erhöhte Reflexivität
- B.3.4 Autopsychische Depersonalisation
- C.1.2 Zwangähnliches Perseverieren zurückliegender Vorgänge
- C.2.3 s12 Abnorm langes Haften optischer Reize, nachträgliches Sehen
- C.2.5 s2 Abnorm langes Haften akustischer Reize, nachträgliches Hören
- C.2.11 Derealisation
- D.1 s1 Somatopsychische Depersonalisation

Kognitive Einbußen (Cluster BS5)

- C.1.8 Störungen des unmittelbaren Behaltens
- C.1.9 Störungen des Kurzzeitgedächtnisses
- C.1.10 Besonders strukturierte Störungen des Langzeitgedächtnisses
- C.1.12 Verlangsamung und Erschwerung der Denkvorgänge
- C.1.13 Störung der Denkinitiative und gedanklichen Intentionalität

Kognitive Störungen (Cluster BS6)

- A.8.4 Unfähigkeit, die Aufmerksamkeit zu spalten
- C.1.1 Gedankeninterferenz
- C.1.3 Gedankendrängen, Gedankenjagen
- C.1.4 Blockierung des jeweiligen Gedankenganges
- C.1.6 Störung der rezeptiven Sprache
- C.1.7 Störung der expressiven Sprache
- C.1.16 Störung der Symbolerfassung (Konkretismus)
- C.1.17 'Subjekt-Zentrismus' - Eigenbeziehungstendenz
- C.2.9 Fesselung (Bannung) durch Wahrnehmungsdetails

Störungen der Körperwahrnehmung (Cluster BS7)

- C.2.3 s6 Wahrnehmungsveränderungen am eigenen Gesicht
- D.1 Taubheits- und Steifigkeitsempfindungen
- D.2 Sensationen motorischer Schwäche ('Lähmungssensationen')
- D.3 Mehr umschriebene Schmerzempfindungen
- D.4 Wandersensationen
- D.5 Elektrisierungssensationen
- D.6 Thermische Sensationen
- D.7 Bewegungs-, Zug- und Druckempfindungen im Körperinneren / an der Körperoberfläche
- D.8 Sensationen abnormer Leere, Leichtigkeit, Schwere, Fall- und Sink-, Levitations- und Elevationsphänomene
- D.9 Sensationen der Verkleinerung, Schrumpfung, Einschnürung, Vergrößerung, Ausdehnung
- D.10 Kinästhetische Sensationen
- D.11 Vestibuläre Sensationen
- D.14 Dysästhetische Krisen

4.2 Vorhersagemodelle

4.2.1 Diskriminanzanalyse der Einzelitems

Bei der schrittweisen Diskriminanzanalyse über die 20 signifikanten Einzelvariablen (s. Tab. 17) wurde sowohl die Vorwärts-Selektion mit einem Signifikanzniveau von 5% für die Aufnahme als auch die Rückwärts-Elimination mit einem 5%-Signifikanzniveau für den Verbleib gerechnet. Beide Selektionsstrategien führten zu dem gleichen Ergebnis von vier ausgewählten Variablen (s. Tab. 19): 'Zwangähnliches Perseverieren zurückliegender Vorgänge' (C.1.2), 'Gedankendrängen, -jagen' (C.1.3), 'Störung der rezeptiven Sprache' (C.1.6) und 'Wahrnehmungsveränderungen am eigenen Gesicht' (C.2.3 s6).

Diese vier unabhängigen Variablen gingen in die nonparametrische Diskriminanzanalyse nach der k-nearest-neighbour-Methode ein. Sie führte für diese Variablen zu den in Tabelle

20 aufgeführten pooled within-class standardisierten Gruppenmitteln und univariaten Teststatistiken der Varianz.

Tab. 19: Ergebnis der schrittweisen Diskriminanzanalysen der Einzelvariablen (n=120).

BSABS-Item-Nr.	Basissymptom	F-Wert	p (> F)	Wilks' λ	p (< λ)
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	0,017	0,8952	0,5774	0,0001
C.1.1	Gedankeninterferenz	2,395	0,1245	0,6307	0,0001
C.1.2	<i>Zwangähnliches Perseverieren zurückliegender Vorgänge</i>	7,434	0,0074	0,6910	0,0001
C.1.3	<i>Gedankendrängen, -jagen</i>	25,000	0,0001	0,8252	0,0001
C.1.4	Blockierung des jeweiligen Gedankengangs	0,036	0,8494	0,5779	0,0001
C.1.6	<i>Störung der rezeptiven Sprache</i>	14,313	0,0002	0,7352	0,0001
C.1.7	Störung der expressiven Sprache	0,017	0,8953	0,5775	0,0001
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	0,482	0,4891	0,5873	0,0001
C.1.17	'Subjekt-Zentrismus',	3,736	0,0557	0,6513	0,0001
C.2.2 s1	Überempfindlichkeit gegen Licht oder visuelle Reize	0,407	0,5247	0,5831	0,0001
C.2.2 s2	Photopsien	0,012	0,9125	0,5773	0,0001
C.2.3 s5	Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	0,294	0,5886	0,5847	0,0001
C.2.3 s6	<i>Wahrnehmungsveränderungen am eigenen Gesicht</i>	6,998	0,0093	0,6513	0,0001
C.2.5 s1	Veränderung der Intensität / Qualität von Gehörs wahrnehmungen	0,807	0,3710	0,5917	0,0001
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	2,244	0,1369	0,6176	0,0001
C.2.8	Sensorische Überwachheit	0,243	0,6230	0,5792	0,0001
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	0,304	0,5827	0,5809	0,0001
C.2.11	Derealisation	0,031	0,8615	0,5777	0,0001
C.3.1	Motorische Interferenz	1,583	0,2110	0,6054	0,0001
C.3.3	Automatismenverlust	0,970	0,3268	0,5969	0,0001

Kursiv: Statistische Kennwerte für den Einschluß bei Vorwärts-Selektion
Standard: Statistische Kennwerte für den Ausschluß bei Rückwärts-Elimination

Tab. 20: Nonparametrische Diskriminanzanalyse der vier Einzelitems: pooled within-class standardisierte Gruppenmittel und Ergebnisse der univariaten Teststatistik.

	Gruppenmittel		F-Statistik, df = 1			
	G ₀	G ₁	pooled s	Korrelation	F	p
C.1.2	- 0,2510	0,1267	0,4060	0,0704	8,9372	0,0034
C.1.3	- 0,4199	0,4962	0,3664	0,1748	25,0000	0,0001
C.1.6	- 4,4093	0,4837	0,3790	0,1676	23,7535	0,0001
C.2.3 s6	- 0,2448	0,2893	0,3221	0,0672	8,4972	0,0043

G₀: Gruppe der Patienten ohne Entwicklung einer schizophrenen Störung im Katamnese-Intervall

G₁: Gruppe der Patienten mit Entwicklung einer schizophrenen Störung im Katamnese-Intervall

Das Modell erwies sich als sehr hoch signifikant (Wilks' $\lambda = 0,6513$: $F_{\text{exakt}} = 15,3913$, $df = 4$, $p = 0,0001$). Die Rate richtiger Gruppenzuordnungen der Berechnungstichprobe bei Kreuzvalidierung betrug 79,17% (s. Tab. 21).

Tab. 21: Nonparametrische Diskriminanzanalyse der vier Einzelitems: Klassifizierungsleistung und Schätzung der Aposteriori-Wahrscheinlichkeit der Fehlerraten bei Kreuzvalidierung in der Berechnungstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	51	14	65
% von G ₀	78,46%	21,54%	100%
% von n	42,50%	11,67%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	11	44	55
% von G ₁	20,00%	80,00%	100%
% von n	9,17%	36,67%	45,83%
Gesamt	62	58	120
% von n	51,67%	48,33%	100%
Apriori-Wahrscheinlichkeit	0,5000	0,5000	
Aposteriori-Wahrscheinlichkeit			
Stratifiziert	0,1954	0,1881	0,1917
Nicht stratifiziert	0,1563	0,2260	0,1911

Damit war das generierte Modell in der Lage, die Gruppen hoch signifikant zu differenzieren ($2 \times 2 - \chi^2 = 40,7735$, $df = 1$, $p < 0,001$), und führte im McNemar- χ^2 -Test zur Aufdeckung signifikanter Veränderungen zwischen zwei Messungen an einer abhängigen Stichprobe zu keinem signifikanten Unterschied zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung (McNemar- $\chi^2 = 0,2500$, $df = 1$, $p > 0,500$). Da die Summe der falschen Vorhersagen weniger als 30 betrug, wurde hier und auch bei den Vergleichen der anderen Vorhersagemodellen bei der Berechnung des McNemar- χ^2 -Wertes die Kontinuitätskorrektur nach Yates verwandt. Und da hierbei gegen die H_1 , die Annahme, daß ein Unterschied zwischen beobachteter und berechneter Gruppenzugehörigkeit besteht, getestet wird, wurde ein relativ hoher α -Fehler von $\alpha = 25\%$ für die Annahme der H_0 festgelegt, daß keine Veränderung der Gruppenzugehörigkeit stattgefunden hat, da durch die Vergrößerung des α -Fehlers die hier entscheidende Wahrscheinlichkeit des β -Fehlers zumindest indirekt gering gehalten werden kann (Bortz, 1985).

4.2.2 Diskriminanzanalyse der Cluster

Bei der schrittweisen Diskriminanzanalyse über die sieben Cluster (s. Tab. 18) wurde ebenfalls sowohl die Vorwärts-Selektion mit einem Signifikanzniveau von 5% für die Aufnahme als auch die Rückwärts-Elimination mit einem 5%-Signifikanzniveau für den Verbleib gerechnet. Beide Selektionsstrategien führten auch hier zu dem gleichen Ergebnis von einem ausgewählten Cluster (s. Tab. 22): 'kognitive Störungen' (Cluster BS6).

Tab. 22: Ergebnis der schrittweisen Diskriminanzanalysen der Cluster (n=120).

Cluster-Nr.	Cluster	F-Wert	p (> F)	Wilks' λ	p (< λ)
BS1	Überbeanspruchung	2,320	0,1304	0,6895	0,0001
BS2	Dynamische Defizite	0,238	0,6266	0,6599	0,0001
BS3	Wahrnehmungs- und Handlungsstörungen	0,476	0,4916	0,6761	0,0001
BS4	Entfremdung	0,039	0,8429	0,6585	0,0001
BS5	Kognitive Einbußen	1,194	0,2768	0,6668	0,0001
BS6	<i>Kognitive Störungen</i>	<i>53,142</i>	<i>0,0001</i>	<i>0,6895</i>	<i>0,0001</i>
BS7	Störungen der Körperwahrnehmung	1,117	0,2928	0,6733	0,0001

Kursiv: Statistische Kennwerte für den Einschluß bei Vorwärts-Selektion

Standard: Statistische Kennwerte für den Ausschluß bei Rückwärts-Elimination

Dieses Cluster ging als unabhängige Variable mit seinem Summenwert in die nonparametrische Diskriminanzanalyse nach der k-nearest-neighbour-Methode ein. Diese führte zu den in Tabelle 23 aufgeführten pooled within-class standardisierten Gruppenmitteln, die in diesem Fall, da es sich um eine intervallskalierte Variable handelt, exakt den Gruppenzentroiden der kanonischen Diskriminanzanalyse entsprechen.

Tab. 23: Nonparametrische Diskriminanzanalyse des Clusters: pooled within-class standardisierte Gruppenmittel und Ergebnisse der univariaten Teststatistik.

	Gruppenmittel		F-Statistik, df = 1			
	G ₀	G ₁	pooled s	Korrelation	F	p
BS6	- 0,6121	0,7234	1,5477	0,3105	53,1423	0,0001

G₀: Gruppe der Patienten ohne Entwicklung einer schizophrenen Störung im Katamnese-Intervall
 G₁: Gruppe der Patienten mit Entwicklung einer schizophrenen Störung im Katamnese-Intervall

In der Modellprüfung wurde die Gesamtlösung mit einem Wert für Wilks' $\lambda = 0,6895$ sehr hoch signifikant ($F_{\text{exakt}} = 53,1423$, $df = 1$, $p = 0,0001$). Der Anteil zutreffender Gruppenzuordnungen der Berechnungstichprobe bei Kreuzvalidierung betrug 77,50% (s. Tab. 24).

Tab. 24: Nonparametrische Diskriminanzanalyse des Clusters: Klassifizierungsleistung und Schätzung der Aposteriori Wahrscheinlichkeit der Fehlerraten bei Kreuzvalidierung in der Berechnungstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	45	20	65
% von G ₀	69,23%	30,77%	100%
% von n	37,50%	16,67%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	7	48	55
% von G ₁	12,73%	87,27%	100%
% von n	5,83%	40,00%	45,83%
Gesamt	52	68	120
% von n	43,33%	56,67%	100%
Apriori-Wahrscheinlichkeit	0,5000	0,5000	
Aposteriori-Wahrscheinlichkeit			
Stratifiziert	0,3062	0,1285	0,2173
Nicht stratifiziert	0,2666	0,1652	0,2159

Im Gegensatz zu dem Modell der nonparametrischen Diskriminanzanalyse über die Einzelitems war das über die Cluster generierte Modell zwar ebenfalls in der Lage, Gruppenunterschiede hoch signifikant aufzudecken ($2 \times 2\text{-}\chi^2 = 38,7343$, $df = 1$, $p < 0,001$), zeigte jedoch im McNemar- χ^2 -Test bei zweiseitiger Testung eine signifikante Veränderung zwischen beobachteter und berechneter Gruppenzugehörigkeit bereits auf einem 5%-Signifikanzniveau (McNemar- $\chi^2 = 5,7870$, $df = 1$, $p < 0,050$), da für eine große Anzahl von Personen ohne beobachtete Schizophrenieentwicklung durch die Modellfunktion eine schizophrene Störung fälschlicherweise vorausgesagt wurde.

4.2.3 Logistische Regressionanalyse der Einzelitems

Die 20 signifikanten Einzelvariablen (s. Tab. 17) gingen als unabhängige Variablen in die logistischen Regressionsanalysen ein, bei der die beiden Outcome-Gruppen als abhängige Variable fungierten. Die Analysen wurden nach der schrittweisen Wald-Methode mit Zulassung einer Konstanten durchgeführt. Dies erfolgte zum einen mit dem Modus der Vorwärts-Suche mit $p = 0,05$ für den Einschluß und $p = 0,10$ für den Ausschluß und zum anderen mit dem der Rückwärts-Eliminierung mit $p = 0,10$ für den Einschluß und $p = 0,05$ für den Ausschluß. Die beiden unterschiedlichen Selektionsstrategien führten hierbei anders als bei der nonparametrischen Diskriminanzanalyse (vgl. Kap. 4.2.1) zu unterschiedlichen Ergebnissen.

Tab. 25: Ergebnis der schrittweisen logistischen Regressionsanalysen der Einzelvariablen im Vorwärtsmodus ($n=120$).

BSABS-Item-Nr.	Basissymptom	β	σ_β	Wald- χ^2 ($df=1$)	p
C.1.3	Gedankendrängen, -jagen	2,3568	0,7134	10,9140	0,0010
C.1.6	Störung der rezeptiven Sprache	2,0283	0,6486	9,7801	0,0018
C.1.17	'Subjekt-Zentrismus', Eigenbeziehungstendenz	1,4359	0,5960	5,8039	0,0160
C.2.2 s1	Überempfindlichkeit gegenüber Licht oder visuellen Reizen	2,0724	0,8863	5,4673	0,0194
Konstante		- 1,4052	0,3023	21,6135	0,0000

β : Regressionskoeffizient

σ_β : Standardfehler der Regressionskoeffizienten

Während bei der Vorwärts-Methode die Variablen 'Gedankendrängen, -jagen' (C.1.3), 'Störung der rezeptiven Sprache' (C.1.6), 'Subjekt-Zentrismus' (C.1.17) und 'Überempfindlichkeit gegenüber Licht und/oder bestimmten visuellen Wahrnehmungsobjekten' (C.2.2 s1) in die Modellgleichung aufgenommen wurden (Tab. 25), verblieben bei der Rückwärts-Methode die

Variablen 'zwangähnliches Perseverieren zurückliegender Vorgänge' (C.1.2) sowie 'Gedankendrängen, -jagen' (C.1.3), 'Störung der rezeptiven Sprache' (C.1.6) und 'Wahrnehmungsveränderungen am eigenen Gesicht' (C.2.3 s6) in der Gleichung (s. Tab. 26).

Tab. 26: Ergebnis der schrittweisen logistischen Regressionsanalysen der Einzelvariablen im Rückwärtsmodus (n=120).

BSABS-Item-Nr.	Basissymptom	β	σ_β	Wald- χ^2 (df=1)	p
C.1.2	Zwangähnliches Perseverieren zurückliegender Vorgänge	1,2951	0,5550	5,4456	0,0196
C.1.3	Gedankendrängen, -jagen	2,0743	0,7106	8,5210	0,0035
C.1.6	Störung der rezeptiven Sprache	2,2148	0,6421	11,8997	0,0006
C.2.3 s6	Wahrnehmungsveränderungen am eigenen Gesicht	1,7343	0,7535	5,2979	0,0214
Konstante		- 1,4402	0,3093	21,6783	0,0000

β : Regressionskoeffizient

σ_β : Standardfehler der Regressionskoeffizienten

Tab. 27: Schrittweise logistische Regressionsanalyse der Einzelitems, Vorwärtssuche: Klassifizierungsleistung in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	53	12	65
% von G ₀	81,54%	18,46%	100%
% von n	44,17%	10,00%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	11	44	55
% von G ₁	20,00%	80,00%	100%
% von n	9,17%	36,67%	45,83%
Gesamt	64	56	120
% von n	53,33%	46,67%	100%

Während das aus der Vorwärts-Suche resultierende höchst signifikante Regressionsmodell ($\chi^2 = 50,244$, $df = 4$, $p = 0,0000$) zu einer Rate zutreffender Gruppenklassifizierungen von insgesamt 80,83% führte (s. Tab. 27), wurden durch das aus der Rückwärts-Eliminierung ebenfalls höchst signifikante Modell ($\chi^2 = 49,243$, $df = 4$, $p = 0,0000$) nur 76,67% der Fälle

der Berechnungsstichprobe korrekt klassifiziert (s. Tab. 28), so daß das Regressionsmodell mit Rückwärts-Eliminierung zugunsten des Regressionsmodells nach Vorwärts-Selektion verworfen wurde.

Tab. 28: Schrittweise logistische Regressionsanalyse der Einzelitems, Rückwärts-Eliminierung: Klassifizierungsleistung in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	55	10	65
% von G ₀	84,62%	15,38%	100%
% von n	45,83%	8,33%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	18	37	55
% von G ₁	32,73%	67,27%	100%
% von n	15,00%	30,83%	45,83%
Gesamt	73	47	120
% von n	60,83%	39,17%	100%

Das gewählte durch die Vorwärts-Suche generierte Modell konnte Gruppenunterschiede hoch signifikant aufdecken (2×2 - $\chi^2 = 45,3297$, $df = 1$, $p < 0,001$) und führte im McNemar- χ^2 -Test zu keinem signifikanten Unterschied zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung (McNemar- $\chi^2 = 0,0109$, $df = 1$, $p > 0,900$).

4.2.4 Logistische Regressionsanalyse der Cluster

Bei der zweiten Variante der schrittweisen logistischen Regressionsanalyse gingen die sieben Cluster (s. Tab. 18) als unabhängige Variablen in die Berechnungen ein, bei der die beiden Outcome-Gruppen erneut als abhängige Variable fungierten. Die Analysen wurden ebenfalls nach der schrittweisen Wald-Methode mit Zulassung einer Konstanten durchgeführt, zum einen im Vorwärts-Modus mit $p = 0,05$ für den Einschluß und $p = 0,10$ für den Ausschluß und zum anderen im Rückwärts-Modus mit $p = 0,10$ für den Einschluß und $p = 0,05$ für den Ausschluß. Im Gegensatz zu den Analysen der Einzelvariablen (vgl. Kap. 4.2.3) führten hier beide Selektionsmodi zu identischen Regressionsmodellen (s. Tab. 29), die mit einem $\chi^2 = 45,947$ bei einem Freiheitsgrad höchst signifikant wurden ($p = 0,0000$) und

75,83% der Fälle der Berechnungsstichprobe der jeweils richtigen Gruppe zuordneten (s. Tab. 30).

Tab. 29: Ergebnis der schrittweisen logistischen Regressionsanalysen der Cluster in beiden Selektionsmodi (n=120).

Cluster-Nr.	Cluster	β	σ_β	Wald- χ^2 (df=1)	p
BS6	Kognitive Störungen	0,9682	0,1922	25,3727	0,0000
Konstante		- 1,4569	0,3113	21,9056	0,0000

β : Regressionskoeffizient

σ_β : Standardfehler des Regressionskoeffizienten

Tab. 30: Schrittweise logistische Regressionsanalyse der Cluster, Rückwärts-Eliminierung und Vorwärts-Selektion: Klassifizierungsleistung in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G_0	G_1	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G_0)	54	11	65
% von G_0	83,08%	16,92%	100%
% von n	45,00%	9,17%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G_1)	18	37	55
% von G_1	32,73%	67,27%	100%
% von n	15,00%	30,83%	45,83%
Gesamt	72	48	120
% von n	60,00%	40,00%	100%

Das generierte Modell war in der Lage, Gruppenunterschiede hoch signifikant aufzudecken (2×2 - $\chi^2 = 31,4685$; $df = 1$, $p < 0,001$), und zeigte im McNemar- χ^2 -Test einen signifikanten Unterschied zwischen beobachteter und berechneter Gruppenzugehörigkeit bei zweiseitiger Testung nur auf dem 25%-Signifikanzniveau (McNemar- $\chi^2 = 1,5648$, $df = 1$, $p < 0,250$) aufgrund seiner Tendenz zu falsch-negativen Vorhersagen.

4.2.5 Grob-Mengen-Modell

Die Berechnungen nach dem Grob-Mengen-Modell mit Voreinstellungen der 'rule roughness' (0,90) und der 'rule precision' (.10), die eine geringe Anzahl relativ genereller, wenig speziali-

sierter Regeln generieren, führten zu insgesamt 15 deterministischen Entscheidungsregeln auf der Basis von insgesamt sechs unabhängigen Variablen (Tab. 31): ‘Gedankenintereferenz’ (C.1.1), ‘zwangähnliches Perseverieren zurückliegender Vorgänge’ (C.1.2), ‘Gedankendrängen, -jagen’ (C.1.3), ‘Blockierung des jeweiligen Gedankengangs’ (C.1.4), ‘Störung der rezeptiven Sprache’ (C.1.6) und ‘Subjekt-Zentrismus’ (C.1.17).

Tabelle 31: Grob-Mengen-Modell: Vorhersageregeln bei einer ‘rule roughness’ = 0,90 und einer ‘rule precision’ = 0,10 und deren Gütekriterien in absteigender Ordnung der Regelwahrscheinlichkeit (n=120). Die Regeln sind deterministisch, d.h. die einzelnen Ausprägungen der unabhängigen Variablen einer Regel sind mit einem logischen ‘und’ (\cap) verbunden, die Regeln untereinander hingegen mit einem logischen ‘oder’ (\cup).

Vorhersage der Regel für die abhängige Variable ‘Ausgang’	Ausprägung der unabhängigen Variablen in der Entscheidungsregel						p (Regel)	auftretende Fälle (#)
	C.1.1	C.1.2	C.1.3	C.1.4	C.1.6	C.1.17		
1 Keine Schizophrenie	vorh.	vorh.	n. vorh.			n. vorh.	1,000	2
2 Schizophrenie	vorh.		vorh.				0,929	14
3 Schizophrenie			vorh.	vorh.			0,917	12
4 Schizophrenie	vorh.			n. vorh.		n. vorh.	0,909	11
5 Schizophrenie		n. vorh.			vorh.		0,857	21
6 Keine Schizophrenie	n. vorh.		n. vorh.	n. vorh.	n. vorh.	n. vorh.	0,855	55
7 Keine Schizophrenie		n. vorh.	n. vorh.		n. vorh.	n. vorh.	0,847	59
8 Keine Schizophrenie	n. vorh.	n. vorh.		n. vorh.	n. vorh.		0,839	56
9 Schizophrenie		n. vorh.				vorh.	0,824	17
10 Schizophrenie		vorh.				n. vorh.	0,706	17
11 Schizophrenie	n. vorh.	vorh.					0,688	16
12 Keine Schizophrenie	vorh.	vorh.		n. vorh.	n. vorh.	vorh.	0,667	3
13 Keine Schizophrenie			n. vorh.	vorh.	vorh.	n. vorh.	0,600	5
14 Keine Schizophrenie	n. vorh.	vorh.	vorh.	vorh.	n. vorh.		0,500	2
15 Schizophrenie	n. vorh.		n. vorh.	n. vorh.			0,254	67
Max. lokale Stärke	0,34	0,22	0,34	0,21	0,34	0,29		

vorh. = Symptom vorhanden, n. vorh. = Symptom nicht vorhanden

p (Regel): Wahrscheinlichkeit, mit der ein die Entscheidungsregel erfüllendes Objekt der richtigen Gruppe zugeordnet wird:

p(durch die Regel korrekt klassifizierte Fälle | die Regel erfüllende Fälle).

auftretende Fälle (#): Anzahl der Fälle, auf die die Entscheidungsregel angewandt werden kann.

Max. lokale Stärke: Die relative Stärke einer unabhängigen Variablen bei der Unterscheidung der Klassen der abhängigen Variablen mit möglichen Ausprägungen zwischen 0 (schwächste) und 1 (stärkste).

Dreizehn der 15 Regeln trafen mit Wahrscheinlichkeiten von über 60% korrekte Vorhersagen oberhalb des Zufallsniveaus von 50% (s. Tab. 30). Da jedoch bei der Generierung von Re-

geln zur Vorhersage einer Ausprägung der abhängigen Variablen bereits generierte Regeln zur Vorhersage anderer Ausprägungen unberücksichtigt bleiben, können Muster als Entscheidungsregel ermittelt werden, die bereits Teil einer Entscheidungsregel für die Vorhersage einer alternativen Ausprägung der abhängigen Variablen sind, wie dies bei den Regeln 3 bzw. 11 und 14 sowie 6 und 15 der Fall war (s. Tab. 31). Da die Regeln 3 bzw. 11 und 6 deutlich höhere Wahrscheinlichkeiten besaßen als die Regeln 14 und 15, wurden in Fällen, in denen beide zutrafen, letztere zugunsten der ersteren verworfen. Auch in allen anderen Fällen, auf die mehr als eine Entscheidungsregel zutraf, kam die Regel mit der höchsten Regelwahrscheinlichkeit zur Anwendung. Damit ließen sich anhand der generierten Entscheidungsregeln 86,67% der Fälle der Berechnungsstichprobe korrekt klassifizieren (s. Tab. 32).

Tab. 32: Grob-Mengen-Modell: Klassifizierungsleistung der Entscheidungsregeln in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	59	6	65
% von G ₀	90,77%	9,23%	100%
% von n	49,17%	5,00%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	10	45	55
% von G ₁	18,18%	81,82%	100%
% von n	8,33%	37,50%	45,83%
Gesamt	69	51	120
% von n	57,50%	42,50%	100%

Mit Hilfe der aus dem Grob-Mengen-Modell resultierenden Entscheidungsregeln konnten Gruppenunterschiede hoch signifikant aufgedeckt werden (2×2 - $\chi^2 = 64,2334$, $df = 1$, $p < 0,001$). Im McNemar- χ^2 -Test zur Ermittlung signifikanter Veränderungen zwischen zwei Messungen an einer abhängigen Stichprobe zeigte sich kein signifikanter Unterschied zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung auf dem 25%-Signifikanzniveau (McNemar- $\chi^2 = 0,7656$, $df = 1$, $p > 0,250$).

4.2.6 Search Partition Analysis

Die auf dichotome Daten speziell zugeschnittene Variante der rekursiven Partitionierung, die Search Partition Analysis (SPAN), wurde über die 'wichtigsten' 15 unabhängigen Variablen

gerechnet, die jeweils positiv mit dem Ausgang in eine Schizophrenie assoziiert waren und die höchsten Korrelationen mit der abhängigen Variable aufwiesen (s. Tab. 33). Als Effektivitätsmaß wurde die Entropie gewählt, die Voreinstellung für die Straffunktion β war 0, die des 'balancing parameters' γ 0,5. Die Komplexität wurde über die Anzahl der in die Partitionen eingehenden unabhängigen Variablen gemessen, und die Suche erfolgte iterativ mit einer maximalen Verarbeitung von zwei Attributen, unabhängige Variablen oder bereits in einer vorherigen Iteration ermittelte Partition, pro Suchschritt. Als Apriori-Wahrscheinlichkeit wurde die Gruppenverteilung der Berechnungstichprobe zugrunde gelegt.

Tab. 33: Die in die Search Partition Analysis eingehenden 15 unabhängigen Variablen, geordnet nach ihrer 'Wichtigkeit' und ihre Produkt-Moment-Korrelation ρ mit der abhängigen Variable 'Ausgang' (n=120).

BSABS-Item-Nr.	Basissymptom	ρ
C.1.3	Gedankendrängen, -jagen	0,418
C.1.6	Störung der rezeptiven Sprache	0,409
C.1.1	Gedankeninterferenz	0,396
C.1.17	'Subjekt-Zentrismus', Eigenbeziehungstendenz	0,345
C.2.8	Sensorische Überwachheit	0,307
C.1.2	Zwangähnliches Perseverieren zurückliegender Vorgänge	0,265
C.2.2 s2	Photopsien	0,229
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	0,279
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	0,269
C.1.4	Blockierung des jeweiligen Gedankengangs	0,241
C.2.3 s6	Wahrnehmungsveränderungen am eigenen Gesicht	0,259
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensorischem Gebiet	0,250
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	0,271
C.2.3 s5	Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	0,271
C.3.3	Automatismenverlust	0,230

Die Suche konvergierte nach fünf Iterationen. Bei einer Komplexität von $c = 4$ begann die Komplexitätssteigerung ('complexity hull') abzufachen, d.h. eine weiter zunehmende Komplexität mit entsprechend höherer Straffunktion β ergab nur noch eine vergleichsweise geringe Erhöhung der Effektivität. Hier war die gefundene Partition optimal für eine Straffunktion von $0,0055 < \beta < 0,0077$ und lautete:

Ausgang 'Schizophrenie' = (C.1.1 = 1) ∪ (C.1.2 = 1) ∪ (C.1.3 = 1) ∪ (C.1.6 = 1),

wobei 1 = Symptom vorhanden.

Entsprechend ergab sich als komplementärer Term:

Ausgang 'keine Schizophrenie' = (C.1.1 = 0 ∩ C.1.2 = 0 ∩ C.1.3 = 0 ∩ C.1.6 = 0),

wobei 0 = Symptom nicht vorhanden.

Die Effektivität bzw. Reinheit der Partition war hoch, die an der Straffunktion gewichtete Entropie, die ein Maß für die Nicht-Reinheit der Partition darstellt, lag bei einem Wert von nur 0,1033. Die ermittelte Partition klassifizierte insgesamt 80,83% der Fälle in der Berechnungsstichprobe korrekt (s. Tab. 34).

Tab. 34: Partitionsmodell: Klassifizierungsleistung in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	51	14	65
% von G ₀	78,46%	21,54%	100%
% von n	42,50%	11,67%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	9	46	55
% von G ₁	16,36%	83,64%	100%
% von n	7,50%	38,33%	45,83%
Gesamt	60	60	120
% von n	50,00%	50,00%	100%

Damit war die Search Partition Analysis in der Lage, Gruppenunterschiede hoch signifikant aufzudecken, und das Modell damit signifikant ($2 \times 2 - \chi^2 = 45,9525$, $df = 1$, $p < 0,001$). Darüber hinaus führte die durch das Partitionsmodell vorgenommene Klassifizierung im McNemar- χ^2 -Test zu keinem signifikanten Unterschied zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung auf dem 25%-Signifikanzniveau (McNemar- $\chi^2 = 0,8804$, $df = 1$, $p > 0,250$).

4.2.7 Diagnostische Effizienzindizes

Insgesamt acht Basissymptome waren bei mindestens 25% der später Schizophrenen bei der Indexuntersuchung vorhanden und besaßen darüber hinaus eine positive prädiktive

Stärke von PPP $\geq 0,65$ sowie einen Prozentsatz falsch-positiven Vorhersagen von weniger als 10% (s. Tab. 35; für alle Einzelitems vgl. Anhang I): 'Unfähigkeit, die Aufmerksamkeit zu spalten' (A.8.4), 'Gedankeninterferenz' (C.1.1), 'zwangähnliches Perseverieren zurückliegender Vorgänge' (C.1.2), 'Gedankendrängen, -jagen' (C.1.3), 'Blockierung des jeweiligen Gedankengangs' (C.1.4), 'Störung der rezeptiven Sprache' (C.1.6), 'Subjekt-Zentrismus' (C.1.17) und 'Derealisation' (C.2.11).

Tab. 35: Diagnostische Effizienzindizes der Basissymptome mit einer Sensitivität $\geq 0,25$, einer positiven prädiktiven Stärke $\geq 0,65$ und einem Prozentsatz falsch-positiver Vorhersagen $\leq 10\%$ in der Berechnungsstichprobe (n=120).

BSABS-Item-Nr.	Sensitivität	Spezifität	PPP	NPP	%F-P	%F-N
A.8.4	0,26	0,94	0,78	0,60	3,3	34,2
C.1.1	0,44	0,91	0,80	0,66	5,0	25,8
C.1.2	0,35	0,88	0,71	0,61	6,7	30,0
C.1.3	0,38	0,95	0,88	0,65	2,5	28,3
C.1.4	0,36	0,86	0,69	0,62	7,5	29,2
C.1.6	0,40	0,94	0,85	0,65	3,3	27,5
C.1.17	0,38	0,91	0,78	0,63	5,0	28,3
C.2.11	0,26	0,89	0,67	0,59	5,8	34,2

PPP: positive prädiktive Stärke
 NPP: negative prädiktive Stärke
 %F-P: Prozentsatz falsch-positiver Vorhersagen
 %F-N: Prozentsatz falsch-negativer Vorhersagen

Diese acht die Selektionskriterien erfüllenden Einzelvariablen erbrachten die beste Klassifizierungsrate (81,67%) bei einem cut-off-Wert von ≥ 2 , d.h. wurde bei dem Vorliegen von mindestens zwei dieser acht Basissymptome eine Schizophrenie prognostiziert und bei einem Vorliegen von weniger als zwei der acht Basissymptome keine Schizophrenie vorhergesagt, so wurden 81,67% der Patienten der Berechnungsstichprobe korrekt zugeordnet (s. Tab. 36).

Bei einem cut-off ≥ 1 betrug die Klassifizierungsrate nur 71,67% und bei ≥ 3 immerhin noch 75,83%, wobei sie bei höheren cut-offs rapide weiter absank. Somit konnten bei dem gewählten cut-off Gruppenunterschiede hoch signifikant aufgedeckt werden ($2 \times 2 - \chi^2 = 47,6417$, $df = 1$, $p < 0,001$). Im McNemar- χ^2 -Test wurde kein signifikanter Unterschied zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung deutlich (McNemar- $\chi^2 = 0,1023$, $df = 1$, $p > 0,500$).

Tab. 36: Variablen mit einer Sensitivität $\geq 0,25$ und einer positiven prädiktiven Stärke $\geq 0,65$: Klassifizierungsleistung bei einem cut-off ≥ 2 in der Berechnungstichprobe ($n = 120$).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	55	10	65
% von G ₀	84,62%	15,38%	100%
% von n	45,83%	8,33%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	12	43	55
% von G ₁	21,82%	78,18%	100%
% von n	10,00%	35,83%	45,83%
Gesamt	67	53	120
% von n	55,83%	44,17%	100%

4.2.8 Diagnostische Likelihood Ratios

Insgesamt 17 Einzelvariablen zeigten eine positive diagnostische Likelihood Ratio von mindestens vier (s. Tab. 37; für alle Einzelitems s. Anhang I). Diese 17 Einzelitems erbrachten die beste allgemeine Klassifizierungsrate (78,33%) bei einem cut-off-Wert von ≥ 3 , wobei allerdings die Vorhersageleistung für eine ausbleibende Entwicklung einer schizophrenen Störung (93,85%) klar über der einer späteren Schizophrenieentwicklung (60,00%) lag (Tab. 38). Bei einem cut-off ≥ 2 war dieses Verhältnis zwar deutlich ausgeglichener (richtige Vorhersage 'keine Schizophrenie' = 78,46%; richtige Vorhersage 'Schizophrenie' = 74,55%), die generelle korrekte Klassifizierungsrate betrug jedoch lediglich 76,67%, so daß entsprechend der Versuchsplanung die Wahl auf den cut-off ≥ 3 fiel. Bei einem cut-off ≥ 4 lag der Prozentsatz richtiger Gruppenzuordnungen nur noch bei 70,83% und sank bei höheren cut-off-Werten weiter ab.

Mit dem für die das Selektionskriterium erfüllenden Einzelitems eingeführten cut-off ließen sich Gruppenunterschiede hoch signifikant aufdecken ($2 \times 2 - \chi^2 = 40,5030$, $df = 1$, $p < 0,001$), damit war das Modell hoch signifikant. Jedoch zeigte sich im McNemar- χ^2 -Test wie auch schon bei der nonparametrischen Diskriminanzanalyse sowie der logistischen Regressionsanalyse der Cluster (vgl. Kap. 4.2.2 und 4.2.4) eine signifikante Veränderung zwischen der beobachteten und der berechneten Gruppenzugehörigkeit bei zweiseitiger Testung (McNemar- $\chi^2 = 11,7789$, $df = 1$, $p < 0,001$) bereits bei $\alpha = 0,001$, die vor allem auf die große Zahl falsch-negativer Schizophrenievorhersagen zurückzuführen war.

Tab. 37: Diagnostische Likelihood Ratios der Basissymptome mit einer positiven diagnostischen Likelihood Ratio $\geq 4,00$.

BSABS-Item-Nr.	Basissymptom	DLR ⁺	DLR ⁻
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	4,33	0,79
C.1.1	Gedankeninterferenz	4,89	0,62
C.1.3	Gedankendrängen, -jagen	7,60	0,65
C.1.6	Störung der rezeptiven Sprache	6,67	0,64
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	4,80	0,80
C.1.17	'Subjekt-Zentrismus', Eigenbeziehungstendenz	4,22	0,68
C.2.2 s1	Überempfindlichkeit gegenüber Licht oder visuelle Reize	7,33	0,80
C.2.2 s2	Photopsien	5,33	0,87
C.2.3 s1	Nah-, Fernsehen	4,00	0,97
C.2.3 s2	Mikro- und Makropsien	6,00	0,95
C.2.3 s6	Wahrnehmungsveränderungen am eigenen Gesicht	4,40	0,82
C.2.3 s7	Scheinbewegungen von Wahrnehmungsobjekten	6,00	0,95
C.2.3 s9	Störungen der Schätzung von Entfernungen, Größe von Gegenständen	4,00	0,97
C.2.3 s12	Abnorm langes Haften optischer Reize, nachträgliches Sehen	4,00	0,97
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	4,00	0,81
C.2.8	Sensorische Überwachheit	20,00	0,81
C.3.3	Automatismenverlust	4,40	0,82

DLR⁺: positive diagnostische Likelihood Ratio; DLR⁻: negative diagnostische Likelihood Ratio

Tab. 38: Einzelitems mit einer positiven diagnostischen Likelihood Ratio ≥ 4 : Klassifizierungsleistung bei einem optimalen cut-off von ≥ 3 in der Berechnungsstichprobe (n=120).

Beobachtete Gruppenzugehörigkeit	Vorhergesagte Gruppenzugehörigkeit		Gesamt
	G ₀	G ₁	
Gruppe ohne Übergang in eine schizophrene Erkrankung (G ₀)	61	4	65
% von G ₀	93,85%	6,15%	100%
% von n	50,83%	3,33%	54,17%
Gruppe mit Übergang in eine schizophrene Erkrankung (G ₁)	22	33	55
% von G ₁	40,00%	60,00%	100%
% von n	18,33%	27,50%	45,83%
Gesamt	83	37	120
% von n	69,17%	30,83%	100%

4.2.9 Vergleich der ausgewählten Variablen

Die generierten Vorhersagemodelle griffen somit auf insgesamt 23 der 72 Einzelitems zurück, 13 von ihnen waren dabei nur in jeweils einem Modell bzw. nur durch die Auswahl ihres Clusters berücksichtigt worden (für eine Übersicht der in die Modelle eingegangenen Variablen s. Anhang J). Nur sechs der aufgenommenen Variablen, die 'Störung der Symbolerfassung' (C.1.16) als Bestandteil des Clusters 'Kognitive Störungen' (BS6), die anderen ausnahmslos optische Wahrnehmungsveränderungen (C.2.3 s1, s2, s7, s9, s12; s. Tab. 37), waren nicht unter den Einzelvariablen mit signifikanten Häufigkeitsunterschieden in den beiden Gruppen (s. Tab. 17) zu finden, während andererseits nur drei dieser signifikanten Einzelvariablen nicht in den Modellen berücksichtigt wurden: 'Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer' (C.2.3 s5), 'Veränderung der Intensität/Qualität von Gehörswahrnehmungen' (C.2.5 s1) und 'motorisch Interferenz' (C.3.1). Die fünf in den χ^2 -Tests nicht signifikant gewordenen optischen Wahrnehmungsveränderungen waren allesamt ausschließlich Bestandteil des Selektionsmodells nach den diagnostischen Likelihood Ratios (s. Tab. 37, vgl. Anhang J), das mit seinen 17 Variablen nicht nur die bei weitem größte Anzahl von unabhängigen Variablen für die Vorhersage der Zielvariable nutzte, sondern mit zehn von ihnen, die einzig in dieses Modell eingehen, auch die größte Variation in die gesamte Variablenauswahl einbrachte.

In jedem der acht generierten Vorhersagemodelle fanden hingegen das 'Gedankendrängen, -jagen' (C.1.3) und die 'Störung der rezeptiven Sprache' (C.1.6) Berücksichtigung, die 'Gedankeninterferenz' (C.1.1) und der 'Subjektzentrismus' (C.1.17) immerhin noch in sechs der acht und die 'Unfähigkeit, die Aufmerksamkeit adäquat zu spalten' (A.8.4), das 'zwangähnliche Perseverieren zurückliegender Vorgänge' (C.1.2) und die 'Gedankenblockierungen' (C.1.4) noch in vier der acht Modelle (vgl. Anhang J). Ein Vorhersagemodell, das sich ausschließlich auf diese sieben häufig ausgewählten Variablen stützt, fand sich nicht.

4.3 Klassifizierungsleistung der Vorhersagemodelle

Zur besseren Abschätzung der Klassifizierungsleistung der Vorhersagemodelle wurden eine Überprüfung der jeweiligen Lösungen anhand einer Validierungsstichprobe von 40 Patienten vorgenommen, die aus der Gesamtstichprobe von $N = 160$ zuvor randomisiert ausgewählt worden und bei der Berechnung der Modelle unberücksichtigt geblieben war. Die Lösungen der einzelnen Methoden führten zu den in Tabelle 39 aufgeführten Gruppenzuordnungen.

Tab. 39: Klassifizierungsleistung der mit den unterschiedlichen Methoden generierten Vorhersageregeln in der Validierungsstichprobe (n = 40).

Methode	Vorhersagen für Patienten ohne Schizophrenieentwicklung (G ₀); n = 16: Anzahl und Anteil von G ₀		Vorhersagen für Patienten mit Schizophrenieentwicklung (G ₁); n = 24: Anzahl und Anteil von G ₁		korrekte Vorhersagen Summe und Gesamtanteil
	G ₀	G ₁	G ₀	G ₁	
	Diskriminanzanalyse der Einzelitems	8 50,00%	8 50,00%	5 20,83%	
Diskriminanzanalyse der Cluster	11 68,75%	5 31,25%	7 29,17%	17 70,83%	28 70,00%
logistische Regression der Einzelitems	10 62,50%	6 37,50%	6 25%	18 75,00%	28 70,00%
logistische Regression der Cluster	14 87,50%	2 12,50%	7 29,17%	17 70,83%	31 77,50%
Analyse nach dem Grob-Mengen-Modell	12 75,00%	4 25,00%	8 33,33%	16 66,67%	28 70,00%
Search Partition Analysis	11 68,75%	5 31,25%	5 20,83%	19 79,17%	30 75,00%
diagnostische Effizienzindizes	11 68,75%	5 31,25%	7 29,17%	17 70,83%	28 70,00%
diagnostische Likelihood Ratios	15 93,75%	1 6,25%	8 33,33%	16 66,67%	31 77,50%

Im Hinblick auf die richtige Zuordnung zu der Gruppe mit einer späteren Entwicklung einer schizophrenen Erkrankung waren die Search Partition Analysis und die Diskriminanzanalyse der Einzelvariablen allen anderen Modellen überlegen und ordneten 79,17% der später schizophren erkrankten Patienten der richtigen Gruppe zu, gefolgt von dem Regressionsmodell der Einzelvariablen mit 75%. Die meisten falsch-negativen Zuordnungen trafen das Grob-Mengen-Modell und das Selektionsmodell über die diagnostischen Likelihood Ratios, die jeweils ein Drittel der späteren Schizophrenen fehlklassifizierten. Die beiden inferenzstatistischen Modelle über die Cluster und das Selektionsmodell über die diagnostischen Effizienzindizes sagten für jeweils 70,83% der in der schizophrenen Gruppe befindlichen Patienten den richtigen Ausgang voraus. Anders stellten sich die Verhältnisse für die korrekten Klassifizierungen in der Gruppe ohne spätere Schizophrenieentwicklung dar. Hier zeigten sich die inferenzstatistischen Modelle über die Einzelitems mit nur 50% korrekter Vorhersagen bei dem diskriminanzanalytischen und 62,50% bei dem Regressionsmodell allen anderen Modellen unterlegen. Das diskriminanzanalytische Clustermodell, das Selektionsmodell über die diagnostischen Effizienzindizes und das Partitionsmodell trafen mit 68,75% richtiger Zuordnungen in dieser Gruppe nur geringfügig präzisere Vorhersagen als das Regressionsmodell der Einzelitems. Überlegen zeigte sich hier das Selektionsmodell über die diagnostischen Li-

likelihood Ratios mit nur einer Fehlklassifizierung und damit 93,75% korrekter Klassifizierungsleistung in der Gruppe ohne spätere Schizophrenieentwicklung, gefolgt von dem Regressionsmodell über die Cluster mit 87,50% richtiger Vorhersagen und damit nur zwei Fehlklassifizierungen. Das Grob-Mengen-Modell war in der Lage, zumindest drei Viertel der Patienten ohne spätere Schizophrenieentwicklung korrekt zu klassifizieren, und nahm damit hinsichtlich der korrekt-negativen Klassifizierungsleistung den dritten Rang in dieser Gruppe ein (s. Tab. 39).

Mit Ausnahme des Modells der kombinierten schrittweisen, nonparametrischen Diskriminanzanalyse über die 20 signifikanten Einzelitems, das zu sehr guten Zuordnungsleistungen in der Gruppe der späteren Schizophrenen geführt hatte, erwiesen sich alle generierten Modelle auch in der Validierungsstichprobe als signifikant (s. Tab. 40).

Tab. 40: Modellprüfungen anhand der Validierungsstichprobe (n = 40).

Methode	2x2-χ^2-Test mit df = 1 bei 2-seitiger Testung zur Aufdeckung signifikanter Gruppenunterschiede durch die Vorhersageregeln(n)		McNemar-χ^2-Test mit df = 1 bei 2-seitiger Testung zur Aufdeckung signifikanter Veränderungen zwischen beobachteter und vorhergesagter Gruppenzugehörigkeit	
	χ^2	p	χ^2	p
Diskriminanzanalyse der Einzelitems	3,72	< 0,100	0,48	> 0,250
Diskriminanzanalyse der Cluster	6,08	< 0,050 *	0,19	> 0,500
logistische Regression der Einzelitems	5,63	< 0,050 *	0,02	> 0,750
logistische Regression der Cluster	13,10	< 0,010 **	2,25	< 0,250 °
Analyse nach dem Grob-Mengen-Modell	6,67	< 0,010 **	1,02	> 0,250
Search Partition Analysis	9,18	< 0,010 **	0,03	> 0,750
diagnostische Effizienzindizes	6,08	< 0,050 *	0,19	> 0,500
diagnostische Likelihood Ratios	14,34	< 0,010 **	4,69	< 0,050 °°°

° signifikanter Unterschied auf dem 25%-Niveau, mit Ablehnung der H_0

°°° signifikanter Unterschied auf dem 5%-Niveau, mit Ablehnung der H_0

* signifikanter Unterschied auf dem 5%-Niveau, mit Annahme der H_1

** signifikanter Unterschied auf dem 1%-Niveau, mit Annahme der H_1

Die beiden insgesamt am besten klassifizierenden Modelle, die schrittweise logistische Regressionsanalyse über die sieben zuvor generierten Cluster und die Selektion nach den po-

sitiven diagnostischen Likelihood Ratios mit Einführung eines cut-off-Wertes, zeigten jedoch signifikante Veränderungen zwischen der beobachteten und der vorhergesagten Gruppenzugehörigkeit (s. Tab. 40), die in beiden Fällen durch eine deutlich höhere Rate fehlklassifizierter Patienten in der Gruppe mit Schizophrenieentwicklung hervorgerufen waren. Beide Modelle tendierten also zu einer besseren Vorhersage des Ausbleibens einer akuten schizophrenen Störung.

Um eine mögliche Überanpassung der Modelle an die Berechnungstichprobe aufzudecken, wurde ein Vergleich der Prozentsätze richtiger und falscher Vorhersagen in der Berechnungs- und Validierungstichprobe gerechnet (s. Tab.41).

Tab. 41: Vergleich der Prozentsätze falscher und richtiger Vorhersagen zwischen Berechnungstichprobe (n = 120) und Validierungstichprobe (n = 40) für die jeweiligen Modelle: Überprüfung einer etwaigen Überanpassung der Modelle an die Berechnungstichprobe.

Methode	Prozentsatz korrekter Vorhersagen in der Berechnungs- und Validierungstichprobe		2x2- χ^2 -Test (df = 1)	
	n = 120	n = 40	χ^2	p
Diskriminanzanalyse der Einzelitems	79,17	67,50	2,7858	< 0,100 *
Diskriminanzanalyse der Cluster	77,50	70,00	1,1622	> 0,250
logistische Regression der Einzelitems	80,83	70,00	2,5304	< 0,250 °
logistische Regression der Cluster	75,83	77,50	0,0624	> 0,750
Analyse nach dem Grob-Mengen-Modell	86,67	70,00	6,5496	< 0,050 **
Search Partition Analysis	80,83	75,00	0,7901	> 0,250
diagnostische Effizienzindizes	81,67	70,00	2,9727	< 0,100 *
diagnostische Likelihood Ratios	78,33	77,50	0,0160	> 0,750

° signifikanter Unterschied auf dem 25%-Niveau, mit Ablehnung der H_0

* signifikanter Unterschied auf dem 10%-Niveau, mit Ablehnung der H_0

** signifikanter Unterschied auf dem 5%-Niveau, mit Ablehnung der H_0

Hierbei zeigten sich signifikante Unterschiede in der Klassifizierungsleistung beider Stichproben und damit eine Überanpassung der generierten Modelle an die Berechnungstichprobe für vier der acht Methoden (s. Tab. 41): schrittweise, nonparametrische Diskriminanzanalyse und schrittweise logistische Regressionsanalyse der Einzelitems, Selektion nach diagnosti-

schen Effizienzindizes mit Einführung eines cut-off-Wertes und die Analyse nach dem Grob-Mengen-Modell. Mit einer Differenz in den Prozentsätzen richtiger Vorhersagen von 16,67% neigte hierbei das Grob-Mengen-Modell am stärksten zu einer Überanpassung an die Berechnungsdaten. Die geringste Anpassung an die Daten der Berechnungsstichprobe fand sich hingegen für das Selektionsmodell über die diagnostischen Likelihood Ratios und das logistische Clustermodell (s. Tab. 41).

Wie erwartet (vgl. Kap. 3.2.6) ergab ein Vergleich der Klassifizierungsleistung der Modelle keinen signifikanten Unterschied in der Anzahl der jeweils durch das einzelne Modell in der Validierungsstichprobe richtig zugeordneten Patienten (1-dimensionales $\chi^2 = 0,5915$, $df = 7$, $p > 0,995$). Und auch keiner der 28 Paarvergleiche der Anzahl richtiger Klassifizierungen durch die mit den acht Methoden generierten Vorhersagemodelle wurde signifikant (2×2 - χ^2 : $0,0582 \leq \chi^2 \leq 1,0031$, $df = 1$, $p > 0,250$).

Die geringfügigen Unterschiede in der Anzahl korrekter Vorhersagen durch die einzelnen Modelle (s. Tab. 39) wiesen dennoch weitgehend in die vorhergesagten Richtungen (vgl. Kap. 3.2.3 bis 3.2.5): Die Variablenselektion anhand der diagnostischen Likelihood Ratios war der anhand der diagnostischen Effizienzindizes in der Klassifizierungsleistung erwartungsgemäß ebenso überlegen wie das Grob-Mengen-Modell der Diskriminanzanalyse der Einzelitems. Hypothesenkonform war zudem, daß sowohl bei der logistischen Regressions- als auch bei der nonparametrischen Diskriminanzanalyse die Cluster-Variante jeweils zu einer höheren Anzahl richtiger Vorhersagen in der Validierungsstichprobe führte als die Einzelvariablen-Variante.

Die Befunde einer gegenüber der logistischen Regression besseren oder zumindest gleichen Klassifizierungsleistung der Vorhersageregeln der Search Partition Analysis ließ sich hingegen nur bedingt bestätigen: Während die logistische Regression der Einzelitems zu einer geringeren Zahl korrekter Zuordnungen als das Partitionsmodell führte, war die der Cluster dem Partitionsmodell überlegen. Nur zum Teil Bestätigung fand auch die Annahme, daß die logistische Regressionsanalyse der nonparametrischen Diskriminanzanalyse generell überlegen sei. Traf dies für das über die Cluster generierte logistische Modell zu, zeigte sich das über die Einzelvariablen berechnete nur dem ebenfalls über die Einzelitems kalkulierten Diskriminanzmodell, nicht aber dem über die Cluster ermittelten Modell überlegen.

Während sich die Vorhersagemodelle hinsichtlich ihrer Klassifizierungsleistung in der Validierungsstichprobe insgesamt gesehen mit Werten zwischen 50 und 93,75% am stärksten in der Anzahl korrekter Vorhersagen des Ausbleibens einer späteren schizophrenen Erkrankung, jedoch kaum in ihrer Gesamtleistung korrekter Klassifizierungen (67,5 bis 77,5%) und in der Anzahl korrekter Vorhersagen späterer schizophrener Erkrankungen (66,67 bis

79,17%) unterschieden (s. Tab. 39), differierten die Vorhersageleistungen hinsichtlich einzelner Patienten deutlich stärker (vgl. Anhang K). Nur 16 der 40 Patienten der Validierungsstichprobe (40%) wurden von allen Vorhersagemodellen korrekt klassifiziert, neun von ihnen hatten im Katamnesezeitraum eine Schizophrenie entwickelt, sieben nicht. Insgesamt sechs Patienten (15%) konnten von keinem der acht Vorhersagemodelle ihrer richtigen Gruppe zugeordnet werden, fünf von ihnen gehörten der Gruppe mit einer späteren Schizophrenieentwicklung an (vgl. Anhang K). Damit fanden sich Unterschiede in der Vorhersageleistung der einzelnen Modelle bei insgesamt 45% der Patienten der Validierungsstichprobe, bei 41,67% der Patienten der Gruppe mit Übergang in eine schizophrene Psychose und bei 50% der Patienten der Gruppe ohne einen solchen.

Die beiden Vorhersagemodelle mit der insgesamt besten Klassifizierungsleistung, das Logitmodell der Cluster und das Selektionsmodell über die diagnostischen Likelihood Ratios, trafen dabei einerseits zwar in den Fällen, in denen fünf bis sieben Modelle zu Fehlzugehörigkeiten führten (Patienten Nr. 37, 54 und 97; vgl. Anhang K), richtige Vorhersagen über die Gruppenzugehörigkeit, andererseits führten sie jedoch in anderen Fällen zu der jeweils einzigen Fehlklassifizierung eines Patienten (Nr. 51, 68 und 160; vgl. Anhang K).

Die bessere Vorhersageleistung einzelner Modelle, insbesondere des logistischen Clustermodells und des Selektionsmodells nach der positiven diagnostischen Likelihood Ratio, gegenüber anderen war damit nicht auf die korrekte Klassifizierung ausschließlich zusätzlicher Patienten zurückzuführen, sondern auf die von Patienten mit anderen Charakteristika bzw. Symptomkonstellationen.

4.4 PLAUSIBILITÄT DER VORHERSAGEMODELLE

Die klinische Plausibilität der generierten Vorhersagemodelle (vgl. Kap. 4.2) wurde von den sieben Ratern recht unterschiedlich beurteilt (s. Abb. 31), was zum einen in der unterschiedlichen Verteilung von Medianen und Modalwerten der Modelle zum Ausdruck kommt, zum anderen in dem Auftreten unterschiedlicher Modalwerte bei einem Vorhersagemodell (s. Abb. 32).

Insgesamt ergaben sich tendenziell positive Beurteilungen der beiden Selektionsmodelle über die diagnostischen Gütekriterien, die diagnostischen Effizienzindizes einerseits (vgl. Kap. 4.2.7) und die positiven diagnostischen Likelihood Ratios andererseits (vgl. Kap. 4.2.8). Beiden Modellen wurde am häufigsten der erste Rang, also die höchste klinische Plausibilität

zugesprochen (s. Abb. 32), wobei dieses Urteil für die Selektion weniger Einzelitems durch die diagnostischen Effizienzindizes eindeutiger ausfiel (s. Abb. 31).

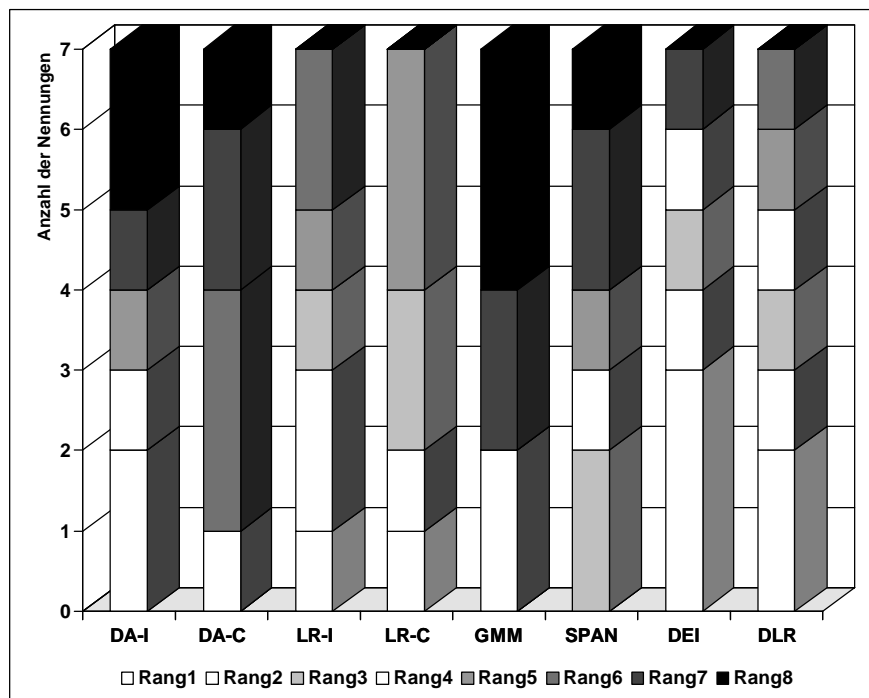


Abb. 31: Klinische Plausibilität der Vorhersagemodelle: Anzahl der den Modellen zugewiesenen Ränge mit Rang 1 = plausibelste Lösung und Rang 8 = am wenigsten plausible Lösung.
 DA-I: diskriminanzanalytisches Modell der Einzelitems
 DA-C: diskriminanzanalytisches Modell der Cluster
 LR-I: logistisches Modell der Einzelitems
 LR-C: logistisches Modell der Cluster
 GMM: Grob-Mengen-Modell
 SPAN: Partitionsmodell
 DEI: Selektionsmodell nach diagnostischen Effizienzindizes
 DLR: Selektionsmodell nach positiven diagnostischen Likelihood Ratios

Relativ homogen war auch die negative Einschätzung der mit dem Grob-Mengen-Modell generierten Entscheidungsregeln (vgl. Kap. 4.2.5), die am häufigsten als aus klinischer Sicht am wenigsten plausibel beurteilt wurden (s. Abb. 31). Die diskriminanzanalytischen und logistischen Modelle (vgl. Kap. 4.2.1 bis 4.2.4), insbesondere die über Einzelitems gerechneten, wurden wie auch das Partitionsmodell (vgl. Kap. 4.2.6) hinsichtlich ihrer klinischen Plausibilität sehr unterschiedlich eingestuft (s. Abb. 31 und 32). Dabei zeigte sich eine Tendenz zu einer positiveren Beurteilung der logistischen Modelle im Vergleich zu dem Partitionsmodell und den diskriminanzanalytischen Modellen.

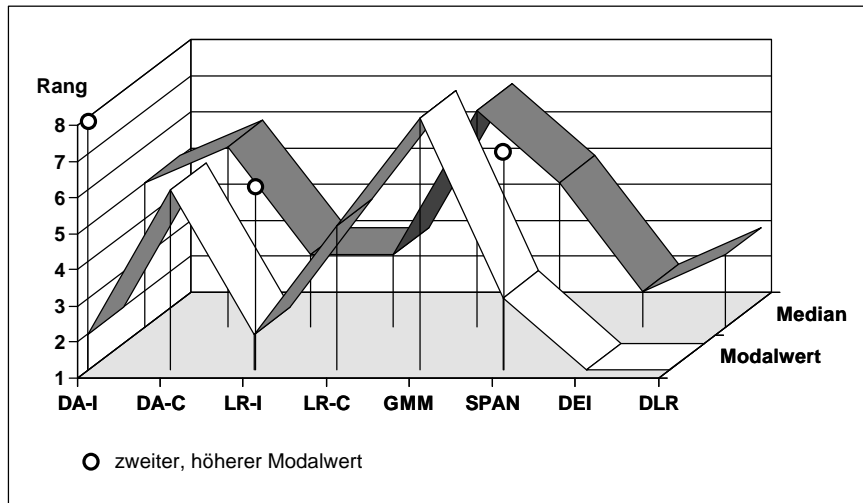


Abb. 32: Klinische Plausibilität der Vorhersagemodelle: Modalwert(e) und Median der für die Modelle vergebenen Ränge.
 DA-I: diskriminanzanalytisches Modell der Einzelitems
 DA-C: diskriminanzanalytisches Modell der Cluster
 LR-I: logistisches Modell der Einzelitems
 LR-C: logistisches Modell der Cluster
 GMM: Grob-Mengen-Modell
 SPAN: Partitionsmodell
 DEI: Selektionsmodell nach diagnostischen Effizienzindizes
 DLR: Selektionsmodell nach positiven diagnostischen Likelihood Ratios

Damit bestätigten sich die aus den in der Literatur ausgeführten theoretischen Überlegungen abgeleiteten Vorhersagen zur klinischen Plausibilität der Prädiktionsmodelle nur zum Teil (vgl. Kap. 3.2.3 und 3.2.4): Die mit dem Grob-Mengen-Modell generierten Entscheidungsregeln wurden nicht nur geringfügig, sondern deutlich als klinisch unplausibler als die durch die Search Partition Analysis generierten eingeschätzt und darüber hinaus auch negativer als die logistischen und diskriminanzanalytischen Modelle bewertet. Hingegen wurden die beiden Selektionsmodelle ähnlich beurteilt, und es zeigte sich eine Tendenz, ein weniger Einzelvariablen umfassendes Modell als klinisch plausibler zu betrachten, in diesem Fall das Selektionsmodell anhand der diagnostischen Effizienzindizes. Insgesamt wurden die Selektionsmodelle entsprechend den Vorannahmen nur geringfügig positiver als die logistischen Modelle und als schon deutlich plausibler als die diskriminanzanalytischen Modelle eingestuft. Das Partitionsmodell, die logistischen Regressionsmodelle und die diskriminanzanalytischen Modelle wurden zwar erwartungsgemäß ähnlich, jedoch sehr heterogen beurteilt. Eine größere Plausibilität des Partitionierungsmodells gegenüber den logistischen und diskriminanzanalytischen Modellen, insbesondere gegenüber dem logistischen Modell mit vorgehaltener Clusteranalyse, zeichnete sich nicht ab.

4.5 ANWENDBARKEIT DER VORHERSAGEMODELLE

Im Gegensatz zu den inhomogenen Einschätzungen der klinischen Plausibilität der Vorhersagemodelle waren die Beurteilungen ihrer Anwendbarkeit übereinstimmender (s. Abb. 33).

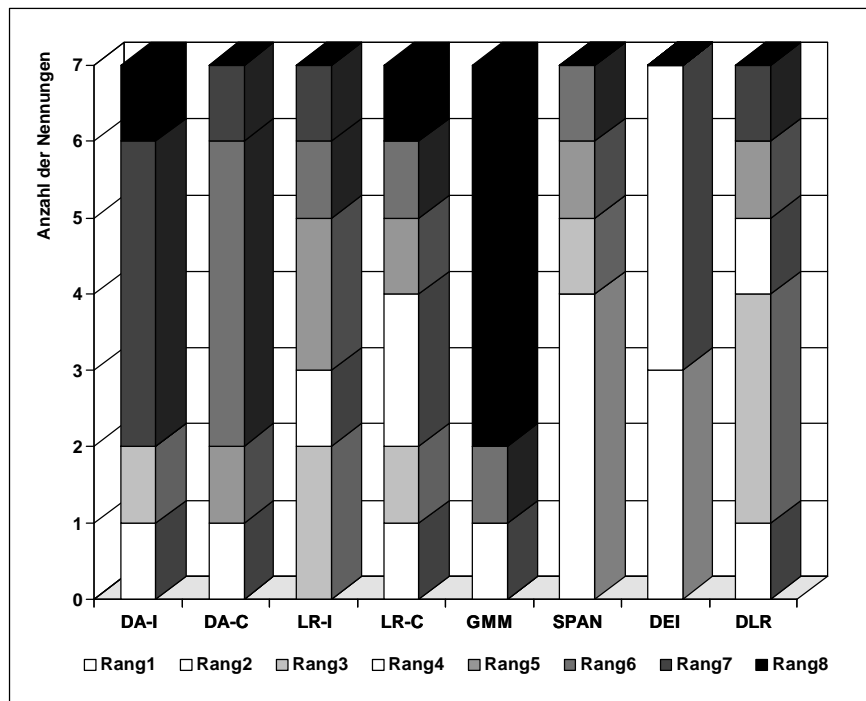


Abb. 33: Anwendbarkeit der Vorhersagemodelle: Anzahl der den Modellen zugewiesenen Ränge mit Rang 1 = erste Wahl und Rang 8 = letzte Wahl.

- DA-I: diskriminanzanalytisches Modell der Einzelitems
- DA-C: diskriminanzanalytisches Modell der Cluster
- LR-I: logistisches Modell der Einzelitems
- LR-C: logistisches Modell der Cluster
- GMM: Grob-Mengen-Modell
- SPAN: Partitionsmodell
- DEI: Selektionsmodell nach diagnostischen Effizienzindizes
- DLR: Selektionsmodell nach positiven diagnostischen Likelihood Ratios

Für jedes Modell wurde ein eigener Rang als Median ermittelt, und dieser stimmte jeweils mit dem Modalwert überein, wobei nur das logistische Regressionsmodell der Einzelitems einen zweiten Modalwert aufwies (s. Abb. 34). Die Vorhersageregeln des Partitionsmodells (vgl. Kap. 4.2.6) wurden gefolgt von den Selektionsmodellen anhand der diagnostischen Effizienzindizes (vgl. Kap. 4.2.7) und anhand der diagnostischen Likelihood Ratios (vgl. Kap. 4.2.8) hinsichtlich ihrer Anwendbarkeit am positivsten beurteilt. Auf den Rängen 4 und 5 befanden sich die logistischen Modelle über die Cluster (vgl. Kap. 4.2.4) bzw. die Einzelitems (vgl. Kap. 4.2.3). Negativer wurde die Anwendbarkeit der diskriminanzanalytischen Modelle beurteilt, wobei das diskriminanzanalytische Modell der Cluster (vgl. Kap. 4.2.2) positiver abschnitt als

das der Einzelitems (vgl. Kap. 4.2.1). Am negativsten wurde die Anwendbarkeit der Entscheidungsregeln des Grob-Mengen-Modells (vgl. Kap. 4.2.5) bewertet.

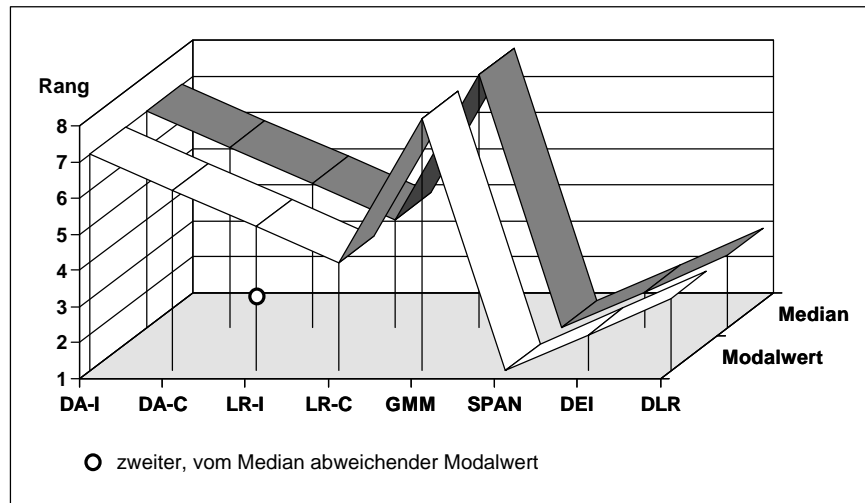


Abb. 34: Anwendbarkeit der Vorhersagemodelle: Modalwert(e) und Median der für die Modelle vergebenen Ränge.

- DA-I: diskriminanzanalytisches Modell der Einzelitems
- DA-C: diskriminanzanalytisches Modell der Cluster
- LR-I: logistisches Modell der Einzelitems
- LR-C: logistisches Modell der Cluster
- GMM: Grob-Mengen-Modell
- SPAN: Partitionsmodell
- DEI: Selektionsmodell nach diagnostischen Effizienzindizes
- DLR: Selektionsmodell nach positiven diagnostischen Likelihood Ratios

Damit stützten auch die Beurteilungen der Anwendbarkeit der generierten Vorhersagemodelle die theoretischen Überlegungen, wie sie hierzu in der Literatur angestellt wurden, ebenfalls nur bedingt: Entgegen den Vorannahmen wurde zwar das Partitionsmodell am anwenderfreundlichsten beurteilt, ihm folgten aber erwartungsgemäß die Selektionsmodelle vor den logistischen Modellen, die ihrerseits positiver als die diskriminanzanalytischen Modelle bewertet wurden. Ebenfalls hypothesenkonform war bei vergleichbaren Regeltypen, den Selektionsmodellen und dem Partitionsmodell in seiner hier ermittelten Form, die Präferenz des jeweiligen Modells mit der geringeren Variablenanzahl und die gegenüber den Selektionmodellen und dem Partitionsmodell negativere Bewertung der Anwendbarkeit des Grob-Mengen-Modells, das wie schon hinsichtlich seiner klinischen Plausibilität auch hinsichtlich der Anwendbarkeit die durchschnittlich schlechteste Bewertung erhielt.

5 DISKUSSION

Standen in der Medizin und insbesondere in der Psychiatrie lange Zeit Bemühungen um eine möglichst einheitliche Diagnostik und adäquate Behandlungsstrategien bestehender Erkrankungen, also ein eher kurativer Ansatz, im Mittelpunkt des Interesses, so wird in den letzten Jahren der Prävention schwerwiegender Erkrankungsbilder immer größere Aufmerksamkeit gewidmet (Klosterkötter, 1998). So war es zwar bereits seit langem unbestritten, daß insbesondere schizophrene Erkrankungen in den seltensten Fällen spontan entstehen, doch wurden die Beschwerdebilder im Vorfeld der akuten Erkrankung als zu unspezifisch für die Diagnose eines schizophrenen Prodroms angesehen:

Besonders bei jungen Menschen tritt vor den typischen schizophrenen Symptomen, [sic] eine Wochen oder Monate dauernde Prodromalphase mit unspezifischen Symptomen auf. ... Diese Symptome sind zwar weder pathognomonisch für eine bestimmte Störung, noch sind sie für die betroffene Person im gesunden Zustand typisch. ... Retrospektiv betrachtet, machen diese Prodromalstadien einen wesentlichen Teil der gesamten Krankheitsentwicklung aus. Es gibt nur wenig Information darüber, ob bei anderen psychiatrischen Störungen ähnliche Prodromi vorkommen, [sic] oder ob ähnliche Zustandsbilder auch zeitweise bei Personen auftreten und wieder zurückgehen, die niemals eine diagnostizierbare psychiatrische Störung entwickeln.

Falls ein Prodrom als typisch und spezifisch für eine Schizophrenie angesehen werden und reliabel beschrieben werden kann und es sich von einem Vorstadium anderer psychiatrischer Störungen oder von einem nicht-krankhaften Zustand eindeutig unterscheidet, kann es gerechtfertigt sein, so ein Prodrom zu den optionalen Kriterien für die Schizophrenie zu zählen. Die zum jetzigen Zeitpunkt vorliegenden Informationen rechtfertigen es allerdings nicht, ein Prodromalstadium als eindeutiges Kriterium für die Diagnose anzusehen. (WHO, 1993, S. 26)

Und genau diese letztgenannte Schwierigkeit zu überwinden und die Basis für eindeutige diagnostische Kriterien für das Prodromalstadium einer Schizophrenie zu schaffen ist vorrangiges Ziel der Früherkennungsforschung auf diesem Gebiet. Um dies aber verlässlich leisten zu können, ist es nicht hinreichend, allein das Augenmerk auf die Erfassung möglicher spezifischer und typischer Prodromalsymptome zu richten und das Studiendesign an die Fragestellung anzupassen, d.h. prospektive Studien durchzuführen, sondern es ist überdies notwendig, Methoden für die Datenanalyse auszuwählen, die in der Lage sind, möglichst optimale valide und reliable Vorhersagemodelle zu generieren.

In der Forschung zum initialen schizophrenen Prodrom wurde bislang jedoch dieser methodische Aspekt zumeist vernachlässigt. So kritisierten Marshall und Kollegen (1994), daß sich nur wenige Studien der Fragestellung widmeten, welche der verfügbaren Techniken die beste und präziseste Vorhersage für einen dichotomen Ausgang liefere, obschon Vorhersagemodelle in der medizinischen Forschung insgesamt unter Klinikern als ein wichtiges Instrument der medizinischen Entscheidungsfindung sehr populär geworden seien und Mathematiker große Anstrengungen unternommen hätten, bestehende Methoden an ein nonparametrisches Skalenniveau anzupassen sowie neue Methoden zur Generierung von Vorhersagemodellen zu entwickeln.

Dieser Frage wurde in der vorliegenden Studie nachgegangen. Acht Vorgehensweisen zur Selektion prädiktiver binärer Variablen und Erstellung von Zuordnungsregeln zu einer von zwei Ausgangsalternativen wurden hinsichtlich ihrer Klassifizierungsleistung in einer Validierungsstichprobe verglichen: schrittweise logistische Regressionsanalysen und kombinierte schrittweise und nonparametrische Diskriminanzanalysen - jeweils mit vorgeschalteter Variablenreduktion zum einen durch Auswahl von Einzelitems nach ihrer statistischen Signifikanz, zum anderen durch eine vorherige hierarchische Clusteranalyse der Items, ein speziell für binäre Daten entwickeltes Partitionierungsverfahren, eine Grob-Mengen-Modell-basierte Analyse und zwei Item-Selektionsverfahren mit Einführung eines Schwellenwertes zur Gruppenzuordnung; hierbei erfolgte die Selektion einerseits nach den diagnostischen Effizienzindizes der Einzelitem, andererseits nach ihren diagnostischen Likelihood Ratios.

Die Analysegrundlage bildeten die Daten von insgesamt 160 Patienten einer prospektiven Studie zur Vorhersage von schizophrenen Störungen anhand von 72 selbstwahrgenommenen subklinischen Beschwerden im Sinne von Basissymptomen (Huber, 1983, 1986; Klosterkötter, Hellmich & Schultze-Lutter, 2000; Klosterkötter, Schultze-Lutter et al., 1997). Harrell und Kollegen (1985) hatten neben der Beachtung und möglicherweise notwendigen Überprüfung der Modellannahmen als Grundvoraussetzung für die Generierung zuverlässiger, klinisch umsetzbarer regressionsanalytischer Vorhersagemodelle angemahnt, daß bereits das Studiendesign der Fragestellung angemessen, d.h. prospektiv, und die Stichprobe ausreichend groß sein müsse, wobei das Zielkriterium ausreichend häufig aufzutreten habe, um das zulässige Verhältnis der Anzahl von Personen mit dem Zielkriterium und der Anzahl der in die Analyse eingehenden unabhängigen Variablen nicht zu überschreiten. Darüber hinaus forderten sie, daß die Nachuntersuchung zur Vermeidung eines Untersuchungsbias möglichst vollständig sein und eine Aufteilung der Daten in eine Berechnungs- und eine Validierungsstichprobe vorgenommen werden müsse, um eine etwaige Überanpassung des Vorhersagemodells an Idiosynkrasien der Berechnungsstichprobe aufdecken zu können (ebd.).

Mit Ausnahme der Forderung nach einer möglichst vollständigen Nachuntersuchung der Stichprobe erfüllen die den Analysen zugrunde liegenden Daten diese Anforderungen, wobei durch die Korrelationsanalysen von möglichen demographischen und anamnestischen Einflußfaktoren und dem Zielkriterium 'Schizophrenie' auch weitgehend sichergestellt wurde, daß die Vorhersage der abhängigen Variable 'Schizophrenie' durch die unabhängigen Variablen 'Basissymptome' nicht auf dritte, konfundierende Variablen zurückzuführen ist. Die Annahme eines direkten Zusammenhangs zwischen dem Vorliegen von Basissymptomen und dem späteren Auftreten einer akuten schizophrenen Psychose wurde überdies durch die bei der Zwischenauswertung der Daten berechnete Kausalanalyse bestätigt (Klosterkötter, Schultze-Lutter et al., 1997). Der einzige signifikante korrelative Zusammenhang zwischen dem Übergang in eine Schizophrenie und anamnestischen Daten ergab sich für die Diagnose einer schizotypen Persönlichkeitsstörung bei der Erstuntersuchung, die auch in der Kausalanalyse einen indirekten Einfluß auf die Entwicklung einer späteren Schizophrenie zeigte. Dieser Zusammenhang ist aber kaum verwunderlich, da die schizotype Persönlichkeitsstörung ohnehin den Schizophrenie-Spektrum-Störungen zugerechnet wird (Olin & Mednick, 1996), in der ICD-10 als Synonym den Begriff der 'prodromalen Schizophrenie' führt (WHO, 1993) und die im DSM-III-R (APA, 1987) aufgeführten Prodromal- bzw. Residualsymptome einer Schizophrenie in der heutigen Fassung des DSM-IV (APA, 1994) weitgehende Überlappungen mit den Diagnosekriterien für die schizotype Persönlichkeitsstörung zeigen. Gerade in Abwesenheit von einer depressiven und/oder ängstlichen Symptomatik wird deshalb das klinische Bild eines initialen schizophrenen Prodroms in vielen Fällen am ehesten den Kriterien der schizotypen Persönlichkeitsstörung entsprechen, so daß diese Diagnose als Folge der prodromalen Symptomatik und nicht als eigenständige und somit in den Kalkulationen zu berücksichtigende Einflußvariable auf die spätere Entwicklung einer schizophrenen Erstmanifestation zu bewerten ist. Über diese bereits bestehenden Hinweise auf einen direkten Zusammenhang von unabhängigen und abhängiger Variable hinaus ist es als ein weiterer Vorteil der vorliegenden Stichprobe zu sehen, daß die Daten der 160 nachuntersuchten Patienten keine fehlenden Werte enthalten, da es Hinweise darauf gibt, daß die unterschiedliche Behandlung fehlender Werte - Eliminierung oder Verrechnung - einen erheblichen Einfluß auf das Ergebnis haben kann (Eisenstein & Alemi, 1996; Marshall et al., 1994).

Ein entscheidender Nachteil der Stichprobe ist allerdings die Tatsache, daß mehr als die Hälfte der Ursprungstichprobe nicht nachuntersucht werden konnte, was bei einem nicht auszuschließenden Bias bei der Selektion der Stichprobe zu einer erheblichen Verzerrung der Vorhersageschätzungen führen kann (Begg & Greenes, 1983; Harrell et al., 1985, Klosterkötter et al., 2001). Da bei der hier vorliegenden methodischen Fragestellung jedoch

nicht auf die Generierung eines validen Vorhersagemodells für die klinische Prognose in Einzelfällen abgezielt wurde, wobei ein solcher Bias bzw. eine Verzerrung der Vorhersageschätzung von erheblicher Bedeutung wäre, sondern auf einen generellen Vergleich der Reliabilität, der Klassifizierungsleistung der unterschiedlichen Vorhersagemodelle, spielt dieser Aspekt nur eine untergeordnete Rolle, da ein solcher Selektionsbias als konstante Störvariable auf alle Vorhersagemodelle gleichermaßen einwirkt. Somit ist nicht zu erwarten, daß er zu einer Verzerrung zugunsten oder zuungunsten bestimmter Modelle führt, die generelle Aussagen über die Klassifizierungsleistung nicht zulässig erscheinen ließe. Zudem hatte sich weder hinsichtlich der Diagnosen, noch des Alters, des Geschlechts oder des Vorliegens von Prodromalsymptomen bei t_1 ein Unterschied zwischen der nachuntersuchten und der nicht nachuntersuchten Gruppe gezeigt, so daß sich keine Hinweise auf einen Selektions-, einen Nachuntersuchungs- oder Verifikationsbias bei t_2 ergaben. Die nachuntersuchte Stichprobe von 160 Patienten kann damit als weitgehend repräsentativ für die Ursprungsstichprobe von 385 Patienten angesehen werden (Klosterkötter et al., 2001).

Bei der Betrachtung der generierten Vorhersagemodelle darf dennoch nicht außer Acht gelassen werden, daß neben der Stichprobenauswahl auch die Auswahl der bei den einzelnen Methoden verfügbaren Kalkulationsoptionen einen erheblichen Einfluß auf das Endmodell hat, etwa die Wahl der Selektionsrichtung und der Ein- und Ausschlußkriterien bei schrittweisen Regressions- und Diskriminanzanalysen, des 'rule precision'- und 'roughness'-Schwellenwertes bei dem Grob-Mengen-Modell, des Komplexitätsmaßes, der Straffunktion und des 'balancing parameters' bei der Search Partition Analyse oder der Auswahlkriterien bei den Selektionsmodellen. So führten etwa in der vorliegenden Studie die Selektionsmodi der logistischen Regressionsanalyse der Einzelvariablen zu weitgehend differierenden Vorhersagemodellen. Darüber hinaus konnte für die schrittweise Regressionsanalyse gezeigt werden, daß sich die Variablenauswahl dramatisch verändert, wenn unterschiedliche Berechnungsstichproben aus der gleichen Grundstichprobe gezogen werden, wobei nicht einmal die 'wichtigsten' Variablen konstant die gleichen blieben (Harrell et al., 1984, 1985).

Idealerweise würden deshalb alle möglichen Stichproben und Optionen berechnet und anschließend ein kalibriertes Modell erstellt werden. Da dies aber bereits bei relativ kleinen Berechnungsstichproben zu einer enormen Zahl von Einzelmodellen führen würde, stellt ein solches Vorgehen bislang nur ein theoretisches Optimum der Modellentwicklung dar - betrachtet man etwa nur die in dieser Arbeit verwandte Stichprobe mit insgesamt 160 Patienten und einer Wahl der randomisierten Berechnungsstichprobe von 120 Personen, so würden sich allein hinsichtlich der möglichen Auswahl der Berechnungsstichprobe schon $(160 \text{ über } 120) = 8,6381\text{E}+37$ Kombinationsmöglichkeiten ergeben und diese Zahl würde noch einmal um den Faktor der möglichen Optionen pro verwandtem Verfahren steigen. Deshalb wurde

in der vorliegenden Arbeit die Schätzung der Anpassung des Modells an die Daten der Berechnungsstichprobe über die Differenz der Vorhersageleistung in der Berechnungs- und Validierungsstichprobe bestimmt und bei der Wahl der Kalkulationsoptionen in Übereinstimmung mit dem Vorgehen von Knuiman, Vu und Segal (1997) auf diejenigen zurückgegriffen, die am ehesten die Routineanwendung der entsprechenden Methode widerspiegeln.

Erwartungsgemäß unterschieden sich die so generierten Vorhersagemodelle hinsichtlich ihrer Klassifizierungsleistung nur geringfügig und wiesen keine signifikanten Unterschiede zwischen den Prozentsätzen richtiger Gruppenzuordnungen auf. Die Richtung der bestehenden Unterschiede in den einzelnen Klassifizierungsleistungen entsprach jedoch weitgehend den in der Literatur beschriebenen Befunden und den daran orientierten Eingangshypothesen.

Die schrittweise logistische Regressionsanalyse über die zuvor ermittelten Variablencluster führte mit 77,5% zu einer Rate an richtigen Klassifizierungen in der Validierungsstichprobe, die nur noch von der Selektionsstrategie nach der positiven diagnostischen Likelihood Ratio erreicht wurde. Beide Verfahren zeigten zudem die geringste Tendenz zu einer Überanpassung der Daten an die Berechnungsstichprobe, neigten jedoch deutlich dazu, das Ausbleiben einer Schizophrenie besser als ihre spätere Entwicklung zu präzisieren.

Die Überlegenheit der schrittweisen logistischen Regressionsanalyse über zuvor generierte Cluster gegenüber fast allen anderen Modellen in der Validierungsstichprobe bestätigt die Befunde des umfassenden Methodenvergleichs von Marshall und Kollegen (1994), die im Vergleich mit der schrittweisen logistischen und logistischen Regressionsanalyse der einzelnen unabhängigen Variablen, der schrittweisen logistischen Regressionsanalyse nach einer vorgeschalteten Hauptkomponentenanalyse, einer auf dem Bayes-Theorem basierenden Analyse, einem additiven Modell, einem subjektiv nach der klinischen Erfahrung bestimmten 'Krankheitswert' und einem hierarchischen Partitionsmodell hierfür ebenfalls die höchste Diskriminationsleistung gemessen an der Fläche unter der ROC-Kurve in der Validierungsstichprobe gefunden hatten. Und auch bei Harrell und Kollegen (1985) hatte das schrittweise logistische Regressionsmodell nach vorheriger Clusteranalyse zu höheren Werten der Fläche unter der ROC-Kurve in der Validierungsstichprobe geführt als das schrittweise Logitmodell der Einzelvariablen und zwei Stratifizierungsverfahren, die rekursive bzw. hierarchische Partitionierung unter Verwendung der CART-Software (Breiman et al., 1984) und das Stratifizierungsverfahren nach der Entropie-Minimax-Methode von Christensen (1981). Allerdings hatte in dieser Studie die schrittweise logistische Regressionsanalyse der Hauptkomponenten die beste Diskriminationsleistung in der Validierungsstichprobe gezeigt (Harrell et al., 1985).

Während sich in beiden Studien die schrittweise logistische Regressionsanalyse zuvor generierter Clustern dem hierarchischen Partitionierungsansatz eindeutig überlegen zeigte, erga-

ben Vergleiche zwischen schrittweiser logistischer Regression von Einzelvariablen und Partitionierungsmodellen ein weniger einheitliches Bild. Während einige Studien eine Überlegenheit des Regressionsmodells vor allem bei umfangreichen Berechnungsstichproben nachwiesen (Hadorn et al., 1992; Knuiman, Vu & Segal, 1997; Long et al., 1993; Marshall et al., 1994; Selker et al., 1995), fanden sich vergleichbare oder gar bessere Klassifizierungsleistungen durch Partitionsmodelle gegenüber logistischen Regressionsmodellen in Studien mit eher kleinen Berechnungsstichproben (Cook & Goldman, 1984; Marshall, 1986; Marshall & Chisholm, 1985; van Mechelen & DeBoeck, 1990). Die Studie von Harrell und Kollegen (1985) verdeutlichte die Abhängigkeit der Klassifizierungsleistung der einzelnen Verfahren von der Größe der Berechnungsstichprobe und der Art der Stratifizierungsmethode. Während das hierarchische mit CART (Breiman et al., 1984) berechnete Partitionsmodell bei der Berechnung über eine kleine Stichprobe ($N = 110$) dem schrittweisen logistischen Regressionsmodell der Einzelvariablen bei der Validierung an einer zweiten Stichprobe leicht überlegen war, verhielt es sich bei einer umfangreicheren Berechnungsstichprobe ($N = 224$) genau umgekehrt. Die Entropie-Minimax-Methode nach Christensen (1981) war hingegen in beiden Stichproben dem logistischen Modell der Einzelvariablen überlegen. Eine Überlegenheit des Partitionierungsansatzes gegenüber dem logistischen Ansatz bei Einzelvariablen bei einer eher kleinen Berechnungsstichprobe von $N = 120$ bestätigte sich erneut in der vorliegenden Studie. Der hier verwandte nicht-hierarchische Partitionierungsansatz, der zu ganz vergleichbaren Ergebnissen wie hierarchische Ansätze führte (Marshall, 1995), die Search Partition Analysis mit der Messung der Entropie als Effektivitätsmaß, konnte in der Validierungsstichprobe 75% der Patienten richtig klassifizieren, die schrittweise logistische Regressionsanalyse der Einzelitems hingegen nur 70%. Dabei war das Partitionsmodell sowohl bei der Vorhersage der späteren Erkrankung als auch bei der ihres Ausbleibens dem logistischen Einzelvariablen-Modell überlegen; die spätere Entwicklung einer schizophrenen Erkrankung wurde mit knapp 80% an korrekten Prognosen in der Validierungsstichprobe von dem Search Partition-Modell sogar gemeinsam mit dem Diskriminanzmodell der Einzelitems am präzisesten vorhergesagt. Damit wies das Partitionsmodell den insgesamt zweitbesten Prozentsatz an korrekten Klassifizierungen auf; hinsichtlich der Vorhersage des Ausbleibens der Erkrankung nahm es eine mittlere Position ein.

Obwohl es Hinweise darauf gibt, daß die Klassifizierungsleistung von hierarchischen Partitionierungsverfahren bei der Validierung an einer unabhängigen Stichprobe stärker abnimmt als die von Regressionsverfahren (Long, Griffith, Selker & D'Agostino, 1993), wies das hier generierte nicht-hierarchische Partitionsmodell im Gegensatz zu dem logistischen Einzelvariablen-Modell keine signifikante Überanpassung an die Berechnungsdaten auf. Die größere Überanpassung herkömmlicher hierarchischer Partitionierungsverfahren resultiert allerdings

daraus, daß sie das Erkrankungsrisiko an immer kleineren einzelnen Substichproben abschätzen, während Regressionsverfahren ihre Risikoabschätzung zu jeder Zeit an der ganzen Stichprobe ausrichten. Dieses Problem der Berechnung über immer kleinere Stichproben wird in der Search Partition Analyse umgangen, indem hier - wie bei inferenzstatistischen Verfahren - die Suche nach der optimalen Partition ebenfalls in jedem Schritt über das Gesamtkollektiv durchgeführt wird, so daß eine lokale Überanpassung an Teilstichproben und damit die Generierung nur lokaler Optima vermieden wird. Dadurch erklärt sich, daß sich auch in der vorliegenden Arbeit keine signifikante Differenz zwischen der Vorhersageleistung in beiden Stichproben für diese speziell für kategoriale Daten entwickelte Partitionierungsmethode fand, die sich damit gerade für klinische Studien mit binären Daten qualifiziert, in denen die Stichprobe nicht frei verfügbar ist, da sie in jedem Schritt die begrenzt verfügbaren Informationen voll auszuschöpfen vermag.

Einen weiteren, vor allem im Hinblick auf die Variablenselektion großen Vorteil der Search Partition Analysis gegenüber den 'klassischen' hierarchischen Partitionierungsverfahren bildet das Abbruchkriterium der iterativen Suche, die dann konvergiert, wenn eine Erhöhung der Komplexität nicht mehr durch eine entsprechende Erhöhung der Partitionsreinheit aufgewogen wird. Die rekursiven Partitionierungsverfahren hingegen setzen die Stratifizierung theoretisch so lange fort, bis jedes Objekt bzw. jede Gruppe identischer Objekte einen eigenen Endpunkt bildet. Abbruchkriterien dieser Verfahren sind zumeist eine willkürlich a priori festgelegte maximale Anzahl von Endpunkten und/oder minimale Zahl von in einem Endknoten enthaltenen Objekten.

Gemeinsam mit dem nicht-hierarchischen Partitionsmodell der Search Partition Analysis erreichte die kombinierte Diskriminanzanalyse der signifikanten Einzelitems hinsichtlich der Prognose einer späteren Erkrankung in der Validierungsstichprobe zwar die besten Klassifizierungsleistungen, wies jedoch insgesamt die geringste Rate an korrekten (67,5%) und richtig-negativen Vorhersagen (50%) und zudem eine signifikante Überanpassung des Modells an die Daten der Berechnungsstichprobe auf, die nur noch von dem Grob-Mengen-Modell übertroffen wurde.

Demgegenüber klassifizierten die kombinierte Diskriminanzanalyse der Cluster, die schrittweise logistische Regressionsanalyse der Einzelitems, das Grob-Mengen-Modell und das Selektionsmodell über die diagnostischen Effizienzindizes immerhin noch jeweils 70% der Patienten der Validierungsstichprobe korrekt. Hierbei zeigte jedoch nur die Diskriminanzanalyse der Cluster keine signifikante Überanpassung des Modells an die Berechnungsstichprobe, die für das Grob-Mengen-Modell, das Diskriminanzmodell der Einzelitems und das Selektionsmodell anhand 'klassischer' diagnostischer Effizienzindizes mit signifikanten Unterschieden zwischen den Prozentsätzen korrekter Klassifizierungen in Berechnungs- und Vali-

dierungsstichprobe von 16,67% und 11,67 % am gravierendsten war. Wie auch in der Vergleichsstudie von Hashemi, Jeloosek und Razzaghi (1993) zeigte sich damit das Grob-Mengen-Modell unter Einbezug der Grenzregion, also unter Nutzung aller verfügbaren Information, dem diskriminanzanalytischen Ansatz über die Einzelitems in Hinblick auf die Gesamtklassifizierungsleistung in der Validierungsstichprobe überlegen. Dies galt auch für die Vorhersageleistung in der Gruppe der nicht schizophren erkrankten Patienten, jedoch nicht in der der später schizophrenen Patienten, für die beide Diskriminanzmodelle dem Grob-Mengen-Modell überlegen waren. Eine solch bessere Vorhersage negativer Ausgänge durch das Grob-Mengen-Modell hatten Hashemi, Jeloosek und Razzaghi (1993) ebenfalls gefunden.

Die im Vergleich mit den anderen Vorhersagemodellen insgesamt jedoch eher geringe Klassifizierungsleistung des Grob-Mengen-Modells in der Validierungsstichprobe und seine ausgeprägte Überanpassung an die Daten der Berechnungsstichprobe sind zum Teil sicherlich erklärbar durch die für methodische Vergleichsstudien relativ kleine Berechnungs- und Validierungsstichprobe bei einer relativ großen Anzahl von Attributen (bei 72 binären Variablen bzw. Symptomen sind dies 144 Kausalattribute bzw. Symptomausprägungen) und der daraus resultierenden enormen Zahl von potentiell möglichen Symptommustern. Da die klinischen Beobachtungen zu Prodromalstadien zudem nahelegen, daß Prodromalsymptome nicht nur theoretisch, sondern auch in der Praxis interindividuell stark variieren und zudem im Verlauf fluktuieren (Birchwood, 1992; Docherty et al., 1978; Jackson et al., 1996; Klosterkötter, 1998; McGlashan, 1996; Vaglum, 1996; WHO, 1993; Yung & McGorry, 1996, 1997), kann angenommen werden, daß diese Heterogenität im vorliegenden Fall dazu geführt hat, daß auch bei einer Wahl der Voreinstellungen für die 'rule roughness' und 'rule precision', die zu der Generierung von eher wenigen, generellen Entscheidungsregeln führt, Artefakte der Berechnungsstichprobe immer noch einen starken Einfluß auf die Lösung nahmen. Dabei beeinflussen solche Artefakte nicht nur die Formulierung der einzelnen Entscheidungsregeln, sondern darüber hinaus auch die Einschätzung der einzelnen Regelwahrscheinlichkeiten und damit die Wahl der anzuwendenden Regel in Fällen, in denen das Symptommuster eines Patienten mehr als eine Entscheidungsregel erfüllt. Unter diesen Gesichtspunkten erstaunt nicht die relativ geringe Klassifizierungsleistung des Grob-Mengen-Modells in der Validierungsstichprobe, sondern vielmehr die Tatsache, daß dieser Ansatz selbst bei einer wie im vorliegenden Fall relativ kleinen Berechnungsstichprobe immer noch zu einer mit inferenzstatistischen Verfahren vergleichbaren oder sogar besseren Vorhersageleistung führt. Bei einer größeren Stichprobe oder weniger komplexen Daten ist somit der Grob-Mengen-basierte Ansatz durchaus als echte Alternative zu den herkömmlichen inferenzstatistischen oder selektiven Ansätzen zu sehen, zumal Grob-Mengen-Modelle die Möglichkeit bieten, aus

'Fehlern' zu lernen, indem sie ihre Regelwahrscheinlichkeiten an neuen Fällen korrigieren und sich so neuen Gegebenheiten anpassen.

Im Vergleich der beiden inferenzstatistischen Verfahren, der schrittweisen logistischen Regressionsanalyse und der kombinierten schrittweisen und nonparametrischen Diskriminanzanalyse, zeigte sich erwartungsgemäß der logistische Regressionsansatz dem diskriminanzanalytischen überlegen (Cook & Goldman, 1984; Jackson, McGorry & Dudgeon, 1995; Marshall et al., 1994). Dies zeigte sich insbesondere in der besseren Klassifizierungsleistung des logistischen Cluster-Modells gegenüber dem diskriminanzanalytischen Cluster-Modell, die ja beide dasselbe Cluster, die kognitiven Störungen, beinhalteten. Mit 7,5% mehr richtigen Klassifizierungen des logistischen Modells, die auf eine bessere Vorhersage negativer Ausgänge zurück ging, zeigte sich deutlich, daß in Stichproben mit nicht-normalverteilten Daten das logistische Regressionsmodell zu verlässlicheren Schätzungen als der diskriminanzanalytische Ansatz führt. Gemeinsam ist jedoch beiden Ansätzen, daß sie im Vergleich zu der Berechnung über die Einzelvariablen insgesamt zu einer besseren Klassifizierungsleistung tendierten, wenn die Berechnung über zuvor generierte Cluster erfolgte. Hierbei wurde die Vorhersage einer spätere Schizophrenieentwicklung zuverlässiger von den jeweiligen Einzelitem-Modellen geleistet, die des Ausbleibens einer Schizophrenie hingegen zuverlässiger von den jeweiligen Cluster-Modellen. Die Cluster-Modelle wiesen zudem eine verglichen mit den Einzelvariablen-Modellen deutlich geringere Tendenz zur Überanpassung der Modelle an die Berechnungsdaten auf und gaben damit eine verlässlichere Schätzung der zu erwartenden Klassifizierungsleistung in der Validierungsstichprobe. Darüber hinaus führte die Berechnung über die Cluster in den jeweiligen Suchoptionen der schrittweisen Diskriminanz- bzw. Regressionsanalyse, Vorwärts-Suche und Rückwärts-Eliminierung, zu den gleichen Schätzfunktionen, wohingegen im Falle der schrittweisen logistischen Regression der Einzelvariablen die beiden Suchoptionen aufgrund der hierbei entstehenden unterschiedlichen partiellen Korrelationskoeffizienten in differierenden Variablenselektionen mit unterschiedlicher Klassifizierungsleistung in der Berechnungsstichprobe resultierten.

Während sich die Klassifizierungsleistungen der ermittelten acht Prognosemodelle mit Prozentsätzen richtiger Vorhersagen von 67,50 bis 77,50% für die gesamte Validierungsstichprobe und mit 66,67 bis 79,19% für die Gruppe der später schizophoren erkrankten Patienten verhältnismäßig gering und mit 50,00 bis 93,75% für die Gruppe der später nicht in eine Schizophrenie übergegangenen Patienten schon offenkundiger unterschieden, differierten ihre Vorhersageleistungen hinsichtlich einzelner Patienten in der Validierungsstichprobe deutlich stärker. So wurde nur für weniger als die Hälfte der 40 Patienten (40,00%) von allen acht Modellen eine zutreffende Prognose hinsichtlich ihres Krankheitsverlaufs, (kein) Übergang in eine schizophrene Psychose, erstellt. Von diesen 16 Patienten hatten neun im Katamnese-

zeitraum eine Schizophrenie entwickelt, sieben hingegen nicht. Insgesamt sechs Patienten der Validierungsstichprobe (15,00%) konnten von keinem der acht Modelle korrekt zugeordnet werden, wobei hier die Gruppe der Patienten mit einem späteren Übergang in eine Schizophrenie deutlich stärker vertreten war und fünf der sechs insgesamt fehlassifizierten Patienten stellte. Damit war bei insgesamt 45% der Patienten die (nicht) korrekte Vorhersage über ihren späteren Krankheitsverlauf abhängig von der gewählten Methode zur Generierung des Vorhersagemodells, bei knapp 42% der Patienten mit Übergang in eine schizophrene Psychose und bei 50% der Patienten ohne Übergang. Diese Zahlen verdeutlichen auf drastische Weise die Abhängigkeit einer zutreffenden Prognose im Einzelfall von der Wahl des Vorhersagemodells, die erheblich größer war, als es bei Betrachtung der Klassifizierungsleistungen der einzelnen Modelle, die insgesamt nur zu einem Unterschied von 10% in der Gesamtklassifizierungsleistung führten, den Anschein hatte.

Aus diesen Zahlen wird auch deutlich, daß besser klassifizierende Verfahren, wie etwa die logistische Regressionsanalyse oder die Variablenselektion nach diagnostischen Likelihood Ratios mit anschließender Einführung eines cut-off-Wertes, nicht etwa nur mehr, sondern auch andere Patienten ihrer jeweiligen Gruppe korrekt zuordnen als Verfahren mit einer schlechteren Klassifizierungsleistung, wie etwa diskriminanzanalytische Ansätze. So trafen abgesehen von dem Partitionsmodell und dem Selektionsmodell der diagnostischen Effizienzindizes alle anderen sechs Verfahren in der Validierungsstichprobe in einem oder zwei Fällen als jeweils einziges Modell eine falsche Vorhersage und dies mit Ausnahme des diskriminanzanalytischen Cluster-Modells immer für Patienten mit einer späteren Schizophrenieentwicklung. Allerdings prognostizierten das Selektionsmodell nach den positiven diagnostischen Likelihood Ratios bzw. dieses Modell und das logistische Cluster-Modell, die hinsichtlich ihrer Gesamtklassifizierungsleistung allen anderen Verfahren überlegen waren, die abhängige Variable auch in den Fällen noch korrekt, in denen (fast) alle anderen Modelle zu einer falschen Vorhersage kamen, wobei es sich hierbei ausnahmslos um Patienten mit negativem Ausgang handelte. Diese bessere Klassifizierungsleistung dieser beiden Modelle für Patienten mit ausbleibender Schizophrenieentwicklung auch in der Einzelfallbetrachtung paßt sich zwar gut in ihr Gesamtbild ein, doch können entsprechende Schlüsse von der allgemeinen Klassifizierungsleistung auf den Einzelfall nicht grundsätzlich gezogen werden. So gehörten im Falle des diskriminanzanalytischen Einzelvariablen-Modells, das mit zu den exaktesten Vorhersagen für einen positiven Ausgang in der Validierungsstichprobe führte, die beiden Patienten, für die es als einziges Modell eine falsche Gruppenzuordnung traf, zu eben dieser Gruppe mit positivem Ausgang.

Die Unterschiede in der individuellen Vorhersageleistung sind wahrscheinlich vor allem darauf zurückzuführen, daß das Stadium der Prodromalphase, in dem sich die übergegangenen

Patienten befanden, nicht bei jedem Patienten das gleiche war und auch nicht ausgeschlossen werden kann, daß sich einige Patienten bei der Index-Untersuchung noch nicht in einem initialen Prodrom, sondern in einem phänomenologisch ähnlichen, aber spontan vollständig remittierenden Vorpostensyndrom (Huber, 1995) befanden. So betrug der Zeitraum zwischen der Index-Untersuchung und der schizophrenen Erstmanifestation in Einzelfällen bis zu 22 Jahren, im Mittel aber nur 5,5 Jahre, was sich gut in Ergebnisse epidemiologischer Untersuchungen einpaßt, die eine mittlere Prodromdauer vor schizophrenen Erstmanifestationen von fünf Jahren ermittelten (Häfner et al., 1995). Darüber hinaus stimmt auch der Prozentsatz von 20% fehlklassifizierten Patienten mit Übergang in eine schizophrene Psychose in der Validierungsstichprobe durchaus mit epidemiologischen Befunden überein, die eine spontane Entwicklung einer schizophrenen Psychose ohne vorheriges initiales Prodrom bei etwa 25% der Patienten mit schizophrener Erstmanifestationen fanden (ebd.). Zukünftige prospektive Langzeitstudien werden zeigen müssen, ob und wie Patienten mit (scheinbar) akutem Krankheitsbeginn im Vorfeld der akuten Psychose erkannt werden können und wie sich unterschiedliche Stadien eines initialen Prodroms einer Psychose mit hinlänglicher Sicherheit differenziert erfassen lassen. Während das erste Problem insbesondere durch eine noch detailliertere Erfassung auch feinsten psychopathologischer Auffälligkeiten zu lösen sein könnte, erfordert das zweite vor allem eine angemessene Analyseverfahren, die anders als die inferenzstatistischen und Selektionsverfahren auch Subgruppen zu identifizieren vermag, wie dies prinzipiell mit Stratifizierungs- und Grob-Mengen-Ansätzen der Fall ist.

Bei der Anwendung jedes mathematischen Verfahrens ist generell jedoch - unabhängig von dem Ergebnis und dessen (scheinbarer) Güte - von entscheidender Bedeutung, daß die der jeweiligen Methode zugrunde liegenden Annahmen bei der Auswahl berücksichtigt werden. So werden in der medizinischen Forschung, in der Kliniker und Statistiker gemeinsam an Projekten arbeiten und die Hypothesengenerierung, Datenerhebung und Datenauswertung zumeist in unterschiedlichen Händen liegen, inferenzstatistische Verfahren, insbesondere (logistische) Regressionsanalysen (Ashby et al., 1992), häufig routinemäßig angewandt, ohne ihre Anwendbarkeit auf die aktuelle Problemstellung zu überprüfen (Appleton, 1995; Harrell et al., 1985), was die Gefahr verzerrter und/oder zu Fehlinterpretationen verleitender Ergebnisse in sich birgt.

Es kann daher als ein Vorteil des Grob-Mengen-Modells als auch von Partitionierungsverfahren betrachtet werden, daß sie auf die Spezifizierung eines immanenten parametrischen Modells verzichten, das den realen Bedingungen möglicherweise nicht gerecht wird. Kritiker dieser Ansätze, vor allem der Partitionierungsansätze, führten jedoch genau dieses Fehlen vorgeschalteter Modellannahmen immer wieder gegen ihre Verwendung ins Feld und wiesen darauf hin, daß hieraus eine Anfälligkeit für das Generieren verfälschter Beziehungen auf-

grund von Stichprobenartefakten resultiere (Doyle, 1973; Einhorn, 1972). Dieser Kritik trägt das hier verwandte Spezialmodell, die Search Partition Analysis, durch seine Grundannahme gerichteter Zusammenhänge Rechnung, was zudem einem Entstehen gemischter Boolescher Funktionen vorbeugt, die dem klinischen Urteil zuwider laufen könnten.

Die Möglichkeit, negative Zusammenhänge aufdecken zu können, kann jedoch dann von Vorteil sein, wenn das Vorliegen einzelner potentiell prädiktiver Variablen in umschriebenen Risikogruppen zu einer Aufhebung der prädiktiven Aussagekraft einer anderen führt, d.h. wenn die empirischen Zusammenhänge komplex und nicht zwangsläufig ausschließlich linearer Natur sind. In solchen Fällen ist zu erwarten, daß das Grob-Mengen-Modell eine bessere Abbildung der Realität liefern wird als auf Additivitätsannahmen beruhende selektive oder inferenzstatistische Verfahren wie etwa Diskriminanz- oder Regressionsanalysen. Die Additivitätsannahme impliziert, daß die Prädiktorvariablen mit dem Zielkriterium prinzipiell in einer additiven Form verbunden sind, wenn das Modell in seiner einfachsten Weise formuliert wird. Bei einem Vorliegen binärer unabhängiger Variablen mit einer 0/1-Kodierung für die Ab- bzw. Anwesenheit der Variable bedeutet dies, daß die den einzelnen vorliegenden, 1-kodierten Variablen zugeschriebenen Gewichte durch das Modell nur aufsummiert werden. Vereinigungen und Schnittmengen von Bedingungen werden hierbei - anders als in Partitionierungs- und Grob-Mengen-Ansätzen - als wenig relevant für die Vorhersage angesehen und statt dessen angenommen, daß ein Patient um so wahrscheinlicher erkrankt, je mehr Symptome er bietet (Harrell et al., 1985). Die logistische Regressionsanalyse läßt zwar auch die Berücksichtigung von Interaktionen in Form von Kreuz-Produkt-Termen zu, doch bietet sie keine Möglichkeit relevante Interaktionen zu identifizieren, ohne daß zuvor alle möglichen Interaktionen als jeweils zusätzliche unabhängige Variable in die Berechnung eingegeben werden. Dies wird jedoch gerade in klinischen Studien mit nicht frei wählbarer Stichprobengröße, in denen die Anzahl der vorliegenden Variablen ihre für die Berechnung zulässige Zahl ohnehin oftmals bereits überschreiten wird, nicht durchführbar sein, so daß bereits im Vorfeld Hinweise auf relevante Interaktionen, wie sie alternative Verfahren bieten können, benötigt werden (Lette et al., 1994).

Die Tatsache, daß für die vorliegenden Daten bei der Search Partition Analysis keine Mehrvariablen-Partitionen gefunden wurden, sowie die vergleichsweise guten Resultate der beiden Selektionsmodelle können als Stützung eines additiven Übergangmodells vom initialen Prodrom in eine akute schizophrene Erstepisode gewertet werden, in dem die Art und Anzahl der vorliegenden Variablen bzw. Symptome eine größere Rolle spielen als etwaige Interaktionen oder Synergismen zwischen ihnen. Dies würde auch die bessere Vorhersageleistung des logistischen Cluster-Modells erklären, da sich in den alternativen Modellen keine eindeutigen Hinweise auf relevante Interaktionen von unabhängigen Variablen ergaben, de-

ren fehlende Berücksichtigung zu einer Minderung der Vorhersageleistung der logistischen Modelle geführt haben könnte, und damit die modellimmanente Additivitätsannahme keine Verletzung der realen Gegebenheiten darzustellen scheint.

Die Möglichkeit, wichtige Interaktionen zwischen einzelnen unabhängigen Variablen aufzudecken und eventuell vorhandene Subgruppen zu identifizieren, die Partitionierungsansätze und Grob-Mengen-Modelle im Gegensatz zu den inferenzstatistischen und Selektionsansätzen bieten, spricht für ihre gemeinsame Anwendung mit regressionsanalytischen bzw. logistischen Verfahren. Die Search Partition Analyse ist zudem in der Lage, über den 'balancing parameter' γ der vorhandenen Randverteilung der abhängigen Variablen in der Berechnungsstichprobe von vornherein Rechnung zu tragen, während etwa für die logistische Regressionsanalyse gezeigt werden konnte, daß sie die besten und zuverlässigsten Ergebnisse nur dann erzielt, wenn die beiden Ausprägungen der abhängigen Variablen in der Berechnungsstichprobe etwa gleiche Häufigkeit besitzen (Harrell et al., 1985), wie dies bei der vorliegenden Stichprobe der Fall war.

Damit boten die Daten der hier verwandten Stichprobe eine optimale Grundlage für die Anwendung einer schrittweisen logistischen Regressionsanalyse (Harrell et al., 1985): Es gab keine Hinweise auf nicht-lineare Beziehungen, auf relevante Interaktionen bestimmter unabhängiger Variablen oder auf konfundierende dritte Variablen, die modellimmanente Annahme einer additiven Beziehung der unabhängigen zu der abhängigen Variable wurde auch durch die Ergebnisse der anderen Verfahren gestützt, die Randverteilung der abhängigen Variable war nahezu balanciert, dem Verhältnis von Fällen mit dem Zielkriterium und den in die Analyse eingehenden Variablen wurde zumindest nach liberalen Anwendungsrichtlinien Rechnung getragen (Backhaus et al., 1987), und nicht zuletzt war das Studiendesign der Fragestellung angemessen, d.h. prospektiv. Vor diesem Hintergrund erscheint die im Vergleich zu fast allen anderen Modellen bessere Vorhersageleistung des logistischen Cluster-Modells wenig erstaunlich und bestätigt die Annahme, daß Regressionsmodelle verglichen mit Stratifizierungsmethoden zu prognostischen Schätzungen von vergleichbarer oder sogar besserer Präzision führen, wenn bei der Analyse die Anwendungsvoraussetzungen hinreichende Beachtung fanden (Harrell et al., 1985).

Hinsichtlich der Auswahl der unabhängigen Variablen in den einzelnen Modellen ergaben sich weitgehende Übereinstimmungen für einen Kernsatz von sieben der insgesamt 23 berücksichtigten Variablen. Die mit 17 aufgenommenen weitaus größte Anzahl von Variablen, von denen zehn ausschließlich in diesem Modell berücksichtigt wurden, fand Eingang in das Selektionsmodell der diagnostischen Likelihood Ratios. Dieses Modell schloß zudem als einziges Variablen ein, die aufgrund ihres geringen Vorkommens nicht in den Häufigkeitsvergleichen der Einzelvariablen signifikant geworden waren, aber durchaus noch befriedigende

positive prädiktive Stärken von $\geq 0,67$ aufwiesen. Diese Tatsache kann als ein Hinweis darauf gewertet werden, daß die anderen der hier verwandten Methoden möglicherweise dazu tendieren, die Wichtigkeit seltener, aber dennoch prädiktiver Variablen bei der Prognose zu unterschätzen.

So testen schrittweise Regressions- und Diskriminanzanalysen den partiellen Einfluß einzelner Variablen auf die Gruppendiskriminierung und tendieren somit zu der Auswahl von Variablen, die auf Einzelitemebene einen deutlichen bzw. signifikanten Häufigkeitsunterschied in den Gruppen aufweisen. Darüber hinaus bergen sie hinsichtlich der Variablenauswahl die Gefahr, daß sich ihre Ergebnisse bei einer größeren Zahl von unabhängigen Variablen aufgrund des Problems multipler Vergleiche nicht replizieren lassen. Dies Problem entspricht dem multipler paarweiser Gruppenvergleiche bzw. dem multipler Vergleiche mehrerer Bedingungsvariablen über zwei Gruppen, bei denen ohne eine entsprechende Fehleradjustierung die Wahrscheinlichkeit, daß einige Vergleiche zufällig signifikant werden, mit steigender Zahl der Vergleiche exponentiell ansteigt.

Bei Regressionsanalysen, bei denen anders als in diskriminanzanalytischen Ansätzen die unabhängige Variablen als fix, die abhängige hingegen als Zufallsvariable behandelt wird, tritt zudem ein weiteres Problem auf: die Regression zum Mittelwert (engl.: „regression to the mean“) und die damit einhergehende sogenannte ‘Schrumpfung’ (engl.: „shrinkage“). Der Effekt der Regression zum Mittelwert impliziert, daß die zukünftigen Werte der abhängigen Variablen dazu tendieren, näher als zu erwarten am generellen Mittelwert zu liegen, also individuelle Unterschiede ‘herausgemittelt’ werden. Ein klassisches Beispiel von Fisher (1958) für diesen Effekt in linearen Regressionsmodellen ist die Vererbung der Körpergröße, wobei gemäß den Vorhersagen der hierbei ermittelten Regressionsgleichung große Väter kleinere Söhne und kleine Väter größere Söhne haben werden. In der linearen Regressionsanalyse ist dieser Effekt bei gleich- und normalverteilten Daten abhängig von der Korrelation der Variablen und um so größer, je kleiner diese Korrelation z.B. zwischen der Größe des Vaters und der des Sohnes ist (Copas, 1997). Eine solche Schrumpfung tritt aber nicht nur bei der linearen Regressionsanalyse auf, sondern wurde auch für logistische Regressionsmodelle gefunden und stellt vor allem bei kleinen Stichproben und/oder einer zu großen Anzahl von unabhängigen Variablen ein ernst zu nehmendes Problem dar (ebd.).

Eine Regression zum Mittelwert besteht prinzipiell nicht bei der Verwendung von Grob-Mengen- und Partitionierungsverfahren und der Selektion nach diagnostischen Gütekriterien, die grundsätzlich seltene, aber hoch prädiktive Zusammenhänge der unabhängigen Variablen mit der abhängigen Variablen erkennen können und nicht a priori dazu tendieren, sie als ‘Ausreißer’ zu eliminieren.

Bei der Selektion nach 'klassischen' diagnostischen Effizienzindizes besteht jedoch das Problem, daß die einzelnen Indizes nicht unabhängig voneinander sind und nur dann insgesamt hohe oder zumindest zufriedenstellende Werte erreichen werden, wenn das Symptom auch in 2×2 - χ^2 -Tests signifikante Häufigkeitsunterschiede in den Gruppen aufweist, da sich die diagnostischen Effizienzindizes wie auch der 2×2 - χ^2 -Test über eine 4-Felder-Tafel ermitteln. Hierdurch besitzen die diagnostischen Effizienzindizes einen starken Bezug zu statistischen Verfahren.

Ein weiteres Problem bei der Selektion nach diagnostischen Effizienzindizes stellt das Fehlen von allgemein gültigen Auswahlkriterien dar und die damit verbundene Unklarheit darüber, welchen Indizes bei der Auswahl das größte Gewicht zukommen sollte (Boyko, 1994). Bei der häufig propagierten alleinigen Betrachtung von Sensitivität und Spezifität unter der Annahme, daß eine hohe Sensitivität eine hohe Erkrankungswahrscheinlichkeit und eine hohe Spezifität eine hohe Wahrscheinlichkeit, nicht zu erkranken, anzeigt (McNeil, Keeler & Adelstein, 1975; Sackett, 1992), kann es zu drastischen Fehlurteilen kommen, da in der Tat auch das Gegenteil zutreffend sein kann (Leisenring & Pepe, 1998). Zudem ist eine Entscheidung zwischen zwei Symptomen oder Tests aufgrund der Sensitivität und Spezifität nur dann eindeutig möglich, wenn sie gleiche Werte für einen der beiden Indizes aufweisen.

Hingegen ist eine Auswahl allein nach positiven und/oder negativen prädiktiven Stärken, die insbesondere in der Früherkennung eine wichtige Rolle spielen, vor allem in Studien problematisch, die nicht repräsentativ für die Normalpopulation sind und einen starken Selektions- und/oder Verifizierung-Bias aufweisen (ebd.). Da beide Indizes von der Prävalenz der Erkrankung in der Berechnungsstichprobe abhängig sind, ist ihre Übertragbarkeit auf andere Populationen in solchen Fällen nicht gewährleistet. Die Berücksichtigung der positiven prädiktiven Stärke bei den hier verwandten Selektionskriterien für die diagnostischen Effizienzindizes stellt damit sehr wahrscheinlich eine Ursache der gefundenen signifikanten Überanpassung des Selektionsmodells an die Daten der Berechnungsstichprobe dar. Dieser Annahme entsprechend wurde eine solche Überanpassung an die Berechnungsdaten nicht für das Selektionsmodell mit alleiniger Berücksichtigung der positiven diagnostischen Likelihood Ratio gefunden, die unabhängig von der Prävalenz der Erkrankung in der Stichprobe ist.

Während also die Tatsache, daß nur Variablen mit signifikanten Häufigkeitsunterschieden in den beiden Ausgangsgruppen Aufnahme in das logistische und diskriminanzanalytische Einzelvariablen-Modell sowie das Selektionsmodell über die diagnostischen Effizienzindizes fanden, aufgrund der ihnen immanenten Selektionsstrategien nicht überrascht, war doch angenommen worden, daß die alternativen Ansätze der Partitionierungs- und Grob-Mengen-Analyse auch Beziehungen und damit Variablen berücksichtigten würden, die in statistische Modelle im weitesten Sinne nicht eingehen. Dennoch wurden auch in diesen Modellen nur

Einzelvariablen, die einen signifikanten Häufigkeitsunterschied zwischen den beiden Ausgangsalternativen aufwiesen, aufgenommen. Dies erklärt sich bei dem Partitionsmodell nach der Search Partition Analysis schon allein aus der Tatsache heraus, daß eine der Vorselektion bei den inferenzstatistischen Ansätzen nach signifikanten Häufigkeitsunterschieden der Einzelvariablen vergleichbare Begrenzung der in die Analyse eingehenden unabhängigen Variablen aufgrund ihrer Produkt-Moment-Korrelation mit der abhängigen Variable erfolgte. Hierbei zeigten naturgemäß nur diejenigen unabhängigen Variablen eine hohe Korrelation mit der abhängigen Variablen, die auch in den Häufigkeitsvergleichen signifikant geworden waren. Demgegenüber ist bei dem Grob-Mengen-Ansatz der Grund für eine ausschließliche Aufnahme von im 4-Felder- χ^2 -Test signifikanten Variablen sicherlich in dem Umstand zu sehen, daß sowohl sehr selten als auch sehr häufig auftretende Variablen insbesondere bei der Generierung weniger, genereller Entscheidungsregeln nur wenig diskriminierende Information für die Mustererkennung beisteuern können, da ihre Ausprägung in den meisten Fällen 'nicht vorhanden', also '0', bzw. 'vorhanden', also '1', sein wird.

Die wichtige Rolle, die Variablen mit signifikanter Häufung in der Schizophrenie-Gruppe zukommt, wird auch durch die Auswahl des Clusters 'Kognitive Störungen' in den inferenzstatistischen Analysen über die zuvor generierten Cluster deutlich: Dieses Cluster beinhaltet mit einer Ausnahme ausschließlich Variablen mit einer solchen Signifikanz, während alle anderen Cluster vorwiegend oder ausschließlich Variablen ohne signifikanten Häufigkeitsunterschied umfassen.

Doch nicht nur vom methodischen Standpunkt aus betrachtet ist die Variablenauswahl plausibel, sondern sie paßt sich darüber hinaus auch inhaltlich gut in die bisherigen Ergebnisse der Basissymptom-Forschung ein: Gerade solche eher eigenartigen kognitiven Denk- und Wahrnehmungsstörungen der Stufe 2, wie sie in der älteren Clusterlösung im BSABS-Subsyndrom 'Informationsverarbeitungsstörungen' zusammengefaßt wurden (Klosterkötter et al., 1996) und hier in die Cluster 'Kognitive Störungen' und 'Wahrnehmungs- und Handlungsstörungen' eingingen, wurden nicht nur vermehrt retrospektiv für die Zeit vor der Erstmanifestation einer schizophrenen Psychose berichtet (Klosterkötter, 1992), sie traten ebenfalls signifikant gehäuft bei Schizophrenen auf in Vergleichen mit psychisch gesunden Personen (Klosterkötter et al., 1996; Klosterkötter, Gross et al., 1997; Schultze-Lutter & Klosterkötter, 1995), mit Patienten mit Störungen durch psychotrope Substanzen, mit neurotischen, Belastungs- und somatoformen Störungen und mit Persönlichkeitsstörungen (Klosterkötter et al., 1996), mit Personen mit einer leichten geistigen Behinderung (Schultze-Lutter & Klosterkötter, 1995) und zumindest in Teilen auch in Vergleichen mit Patienten mit einer affektiven Störung (Bechdolf et al., 1998; Ebel et al., 1989; Klosterkötter et al., 1996) oder einer Adoleszentenkrise (Resch, 1992). Zudem waren Störungen aus diesem Bereich

auch signifikant häufiger bei erstgradigen Verwandten von bereits manifest schizophren erkrankten Personen, die vom Alter her betrachtet noch eine erhöhte Wahrscheinlichkeit für die Entwicklung einer Psychose besaßen (Häfner et al., 1995), als bei nach Alter, Geschlecht und Schulbildung gematchten, nicht familiär belasteten psychisch gesunden Kontrollpersonen aufgetreten, wobei die Häufigkeit dieser kognitiven und perzeptiven Störungen bei Verwandten zwischen der Schizophrener und der gesunder Kontrollen lag (Klosterkötter, Gross et al., 1997). Damit scheint diese Symptomgruppe sowohl state- als auch trait-Charakter zu besitzen, eine Eigenschaft, die ja in besonderem Maße Prodromalsymptomen zugeschrieben wird (Nuechterlein & Dawson, 1984; Parnas, 1999).

Doch nicht nur in der Basissymptomforschung, sondern auch in anderen Entwicklungsmodellen psychotischer Episoden kommt subklinischen Denk- und Wahrnehmungsstörungen besonderes Gewicht zu. So finden sie etwa in dem heute weitgehend anerkannten Vulnerabilitäts-Streß-Bewältigungsmodell (Nuechterlein, 1987; Nuechterlein & Dawson, 1984; Nuechterlein & Zaucha, 1990) sowohl unter den schizotypen Persönlichkeitsmerkmalen als Teil der persönlichen Vulnerabilitätsfaktoren als auch unter den Prodromalsymptomen Berücksichtigung (Klosterkötter, 1996). Und auch in älteren Übergangsmodellen, die zumeist auf der Basis von Beobachtungen von schizophrenen Patienten und deren Selbstbeschreibungen und weniger auf systematischen Erhebungen und Untersuchungen entstanden, werden solche kognitiven und perzeptiven Störungen immer wieder beschrieben: Im Übergangsmodell von Docherty und Mitarbeitern (1978) werden sie als Folge eher adynamischer, noch weitestgehend unspezifischer Symptome als Teil der 'Disinhibition' und der darauf folgenden 'Psychotic Disorganisation' angesehen, während sie im Übergangsmodell von Chapman (1966; McGhie & Chapman, 1961) als primäre, schon spezifische Symptome betrachtet werden, die erst die Grundlage für unspezifische, neurotisch anmutende Symptome sowie akut-psychotische Symptome legen. Damit passen sich die vorliegenden Vorhersagemodelle hinsichtlich ihrer Variablenauswahl inhaltlich gut in die bisherigen Modellvorstellungen und Ergebnisse zu der Entwicklung schizophrener Episoden ein.

Problematischer als die Variablenauswahl stellt sich die subjektive Beurteilung der generierten Vorhersagemodelle durch Experten auf dem Gebiet der Früherkennung hinsichtlich ihrer klinischen Plausibilität und Anwendbarkeit dar. Das Grob-Mengen-Modell wurde sowohl im Hinblick auf seine klinische Plausibilität als auch auf seine Anwendbarkeit von den befragten Experten mit großer Übereinstimmung am negativsten von allen Vorhersagemodellen bewertet. Dies ist sicherlich zum einen dadurch zu erklären, daß die generierten Entscheidungsregeln auf den ersten Blick unübersichtlich und teils widersprüchlich erscheinen, da Teile einer Entscheidungsregel eine eigene Regel mit der Vorhersage eines anderen Ausgangs bilden können. Zum anderen sind in den Entscheidungsregeln für eine Prognose der

Erkrankung auch negative Terme enthalten bzw. in denen für eine Prognose des Ausbleibens der Erkrankung auch positive, was der klinischen Vorannahme, daß das Vorliegen eines Symptoms mit einer größeren Erkrankungswahrscheinlichkeit assoziiert ist, zuwiderläuft. Dies hatte bereits bei der Entwicklung der Search Partition Analyse zu der Restriktion auf die Zulassung ausschließlich regulärer Partitionen geführt, wobei jedoch auch die Abwesenheit einer unabhängigen Variablen als prädiktiv für das Zielkriterium definiert werden kann, wenn sich entsprechende Zusammenhänge in den Korrelationskoeffizienten abzeichnen (Marshall, 1995).

Dieser Verzicht auf gemischte Boolesche Funktionen, die geringe Zahl von nur vier ausgewählten Variablen und die Abwesenheit von Interaktionstermen war vermutlich der Hauptgrund für die positive Beurteilung des Partitionsmodells hinsichtlich seiner Anwendbarkeit mit 'Rang 1' sowohl im Modalwert als auch im Median. Die Vorhersage einer späteren schizophrenen Psychose anhand nur jeweils eines Symptoms führte jedoch wahrscheinlich auch dazu, daß diesem Modell nur ein mittlerer, fünfter Rang für seine klinische Plausibilität zugewiesen wurde, da hiermit ein initiales schizophrenes Prodrom phänomenologisch eindeutiger zu erkennen wäre als eine akute schizophrene Psychose, für die nach den allgemeinen A-Kriterien des DSM-IV (APA, 1994) und nach den ICD-10-Kriterien (WHO, 1993) mindestens ein eindeutiges oder zwei weniger eindeutig ausgeprägte Symptome vorliegen müssen. Die 'Diagnose' eines in der Schizophrenie-Forschung übereinstimmend als wesentlich unspezifischer betrachteten Prodroms auf der Basis nur eines Symptoms erscheint somit vom klinischen Standpunkt nach heutigen Erkenntnissen in der Tat wenig plausibel.

Als wenig klinisch plausibel und ebenfalls wenig anwenderfreundlich wurden neben dem Grob-Mengen-Modell auch die beiden diskriminanzanalytischen Modelle eingeschätzt, die auf dem siebenten bzw. sechsten Rang lagen. Dies ist sicherlich in beiden Fällen auf ihre dimensionale Lösung zurückzuführen, da sich das diskriminanzanalytische Cluster-Modell in seinen eingeschlossenen Variablen nicht von dem logistischen Cluster-Modell unterscheidet, aber dennoch in seiner klinischen Plausibilität drei Ränge unterhalb desselben eingestuft wurde. Die logistischen Modelle ihrerseits wurden ebenfalls insgesamt negativer als die beiden Selektionsmodelle und hinsichtlich ihrer Anwendbarkeit auch negativer als das Partitionsmodell beurteilt. Wie auch bei den diskriminanzanalytischen Modellen wurde hierbei das Cluster-Modell in seiner Anwendbarkeit positiver, in seiner klinischen Plausibilität jedoch negativer als das Einzelvariablen-Modell bewertet. Der Grund hierfür liegt möglicherweise in der Tatsache, daß eine unterschiedliche Gewichtung einzelner Symptome und eine damit einhergehende Rangordnung ihrer 'Wichtigkeit' für den Übergang in eine schizophrene Psychose klinisch sinnvoller erscheint als eine generelle Gewichtung einer Summe von Symptomen, letzteres aber einfacher zu berechnen ist.

Sehr positiv wurden in beiden Aspekten die Selektionsmodelle mit Einführung eines cut-off-Wertes eingestuft, wobei sich die Selektion nach den diagnostischen Effizienzindizes, die zu der Auswahl von acht Variablen führte, der nach diagnostischen Likelihood Ratios mit 17 ausgewählten Variablen hinsichtlich beider Beurteilungsaspekte überlegen zeigte. Beide Modelle wurden in ihrer geschätzten Anwenderfreundlichkeit nur noch von dem Partitionsmodell übertroffen, das in seiner ausformulierten Form jedoch starke Ähnlichkeit mit den Selektionsmodellen aufwies. Hierbei war der zugewiesene mittlere Rang klar von der Anzahl der aufgenommenen Variablen bestimmt, wobei entsprechend den Vorannahmen die Anwendbarkeit um so positiver bewertet wurde, je weniger Variablen das Modell enthielt. Dieser Gesichtspunkt schlug sich vermutlich auch bei der Einschätzung der klinischen Plausibilität nieder, bei der ebenfalls der weniger umfangreichen Selektion nach den diagnostischen Effizienzindizes der Vorzug gegenüber der umfangreicheren nach der positiven diagnostischen Likelihood Ratio gegeben wurde, so daß dem Kriterium von 'mindestens 2 von 8 Symptomen' gegenüber dem von 'mindestens 3 von 17 Symptomen' auch vom klinischen Standpunkt der Vorzug gegeben wurde - wenn auch weniger deutlich als bei der Anwendbarkeit, während das von 'mindestens 1 von 4 Symptomen' als klinisch eher wenig plausibel eingestuft worden war. Die generell positive Beurteilung der Selektionsmodelle ist mutmaßlich zudem auch darauf zurück zu führen, daß sie in ihrer Form der Entscheidungsfindung in der klinischen Praxis am nächsten kommen. So wird im klinischen Alltag die An- bzw. Abwesenheit bestimmter als relevant erachteter Symptome oder Risikofaktoren bei der Prognose oder Diagnose für einen bestimmten Patienten, also eine Strategie des selektiven Ausschlusses (Koss & Feinstein, 1971), eine größere Rolle spielen als eine Gewichtung einer bestimmten vorgegebenen Anzahl von Faktoren, also eine Strategie des graduellen Einschlusses (ebd.). Und auch die heute gültigen Diagnosesysteme, DSM-IV und ICD-10, folgen in der Formulierung ihrer diagnostischen Kriterien der Strategie des selektiven Ausschlusses und tragen somit dazu bei, entsprechende Entscheidungsregeln als klinisch plausibler erscheinen zu lassen als etwa inferenzstatistische Modellgleichungen oder die komplexeren Grob-Mengenbasierten Entscheidungsregeln.

Es darf jedoch nicht außer Acht gelassen werden, daß bei der Bewertung der beiden Selektionsmodelle in ihrer letztendlichen Form ein wichtiger Unterschied zwischen den diagnostischen Effizienzindizes und den diagnostischen Likelihood Ratios nicht zum Tragen kam: Im Gegensatz zu den diagnostischen Likelihood Ratios geben die diagnostischen Effizienzindizes nur Absolutwerte für die Erkrankungswahrscheinlichkeit an, die die Apriori-Erkrankungswahrscheinlichkeit nicht berücksichtigen. Es ist aber vom klinischen Standpunkt aus gesehen kaum vertretbar, anzunehmen, daß etwa das Vorliegen von einem oder mehreren prädiktiven Prodromalsymptomen einer Schizophrenie bei einem Patienten ohne familiä-

re Belastung die gleiche Aussagekraft besitzen solle wie bei einem Patienten mit einer familiären Belastung, da bei letzterem das Erkrankungsrisiko in Abhängigkeit zu seinem Verwandtschaftsgrad mit dem bereits erkrankten Angehörigen von vornherein als erhöht betrachtet würde (Gottesman, 1991) und dem Vorliegen psychischer Auffälligkeiten ein anderes 'Gewicht' bei der klinischen Beurteilung zukäme. Eine solche Verrechnung der Apriori-Erkrankungswahrscheinlichkeit ermöglichen die diagnostischen Likelihood Ratios, die die Zunahme der Apriori-Erkrankungswahrscheinlichkeit bei Vorliegen des Symptoms bzw. ihre Abnahme bei Abwesenheit des Symptoms anzeigen und damit eine Abschätzung des individuellen Erkrankungsrisikos erlauben, ohne daß die entsprechenden bekannten Risikofaktoren in die Berechnung des Vorhersagemodells mit einfließen müssen und möglicherweise durch hohe Interkorrelationen mit prädiktiven Symptomen zu einem Ausschluß eben dieser prädiktiven Symptome führen, wie dies etwa bei der logistischen Regressionsanalyse der Fall sein könnte. Ursprünglich wie die diagnostischen Effizienzindizes als Testgütekriterien konzipiert, können diagnostische Likelihood Ratios jedoch auch für jedes Vorhersagemodell angegeben werden und erlauben so theoretisch eine Berücksichtigung der Apriori-Erkrankungswahrscheinlichkeit für jedes generierte Vorhersage- oder Prognosemodell unabhängig von der Art seiner Entstehung - inferenzstatistische Analyse, Vergleich individueller Datenmuster, Stratifizierung bzw. Partitionierung von Objekten oder Patienten oder Variablenauswahl nach bestimmten Kriterien.

Insgesamt belegt die vorliegende Studie die in anderen methodischen Vergleichsstudien gefundenen Trends für die Klassifizierungsleistung der einzelnen Verfahren und unterstreicht erneut die wichtige Rolle der schrittweisen logistischen Regressionsanalyse bei der Generierung von reduzierten und damit ökonomischen Vorhersagemodellen für eine dichotome Zielvariable, wenn ihren Modellannahmen und Anwendungsvoraussetzungen hinreichend Rechnung getragen wird. Dabei zeigte sich eine eventuell notwendige Vorreduktion der Anzahl der unabhängigen Variablen durch Strukturen-entdeckende Verfahren, wie hier durch eine hierarchische Clusteranalyse, in Übereinstimmung mit bestehenden Methodenvergleichen der Vorselektion von Einzelvariablen anhand von Häufigkeitstests überlegen. Bei der Applikation der logistischen Regressionsanalyse ist jedoch zu bedenken, daß sie dazu tendiert, ihre beste und zuverlässigste Schätzung bei einer balancierten Randverteilung der abhängigen Variablen abzugeben sowie Ausreißer herauszumitteln und damit gerade Extrem- oder Grenzfälle fehlzuklassifizieren, und daß sie darüber hinaus nicht in der Lage ist, nicht-lineare Zusammenhänge, die in Form von Interaktions- bzw. Kreuz-Produkt-Termen einen signifikanten Einfluß auf die abhängige Variable haben, oder Subgruppen selbständig aufzudecken. Deshalb ist es vor allem bei einer Vielzahl von unabhängigen Variablen mit noch nicht hinlänglich gesicherten Erkenntnissen über ihre Beziehungen untereinander und zu der ab-

hängigen Variable und einer voraussichtlich sehr heterogenen Stichprobe mit einem eher niedrigen Anteil von Personen mit dem Zielkriterium, wie dies in Studien zur Früherkennung schizophrener Psychosen meist der Fall sein wird, empfehlenswert, parallel zu der logistischen Regressionsanalyse alternative, nicht an Modellannahmen gebundene Verfahren wie etwa die (nicht-)hierarchische Partitionierung einzusetzen, um Hinweise auf möglicherweise relevante Subgruppierungen oder Variableninteraktionen zu erhalten. Dies sollte insbesondere dann geschehen, wenn neben potentiellen Prodromalsymptomen auch überdauernde Risikofaktoren bzw. Vulnerabilitätsmarker in die Berechnung einfließen, da hierbei zu erwarten ist, daß neben linearen auch anders geartete, etwa exponentielle Zusammenhänge auftreten, die die der logistischen Regressionsanalyse immanente Additivitätsannahme verletzen.

Eine Kombination des logistischen und des Partitionierungsansatzes empfiehlt sich auch für die Datenanalyse von Langzeitstudien mit mehreren Meßzeitpunkten, wobei bei der Auswahl der Modelle solche berücksichtigt werden sollten, die nicht nur den Ausgang an sich, sondern auch den Zeitpunkt seines Eintritts einbeziehen, wie dies etwa das Proportional Hazard Modell von Cox (1972; Harris & Albert, 1991; Kalbfleisch & Prentice, 1980) oder Survival Regression Trees (LeBlanc & Crowley, 1992; Segal, 1988) tun. Für diese Verfahren konnte im Vergleich mit der 'einfachen' logistischen Regression bzw. der 'klassischen' hierarchischen Partitionierung bereits gezeigt werden, daß die volle Nutzung aller über die verschiedenen Meßzeitpunkte erhobenen Informationen anstelle der nur binären Information über den Status zum Zeitpunkt der letzten Erhebung besonders bei langfristig angelegten Studien zu einer Verbesserung der Vorhersageleistung führt (Knuiman, Vu & Segal, 1997).

Bestehen im Vorfeld der Analyse jedoch keine empirischen Hinweise auf oder begründete Vermutungen über die Art der Zusammenhänge zwischen den (un-)abhängigen Variablen, sollte annahmefreien alternativen Verfahren generell der Vorzug gegenüber der logistischen Regression gegeben werden. Für Nominaldaten erzielte hierbei ein auf dieses Datenniveau spezialisierter nicht-hierarchischer Partitionierungsansatz, die Search Partition Analysis (Marshall, 1995), mit der logistischen Regression von zuvor ermittelten Clustern beinahe vergleichbare Ergebnisse in der Gesamtklassifizierungsleistung bei einer Überlegenheit in der Vorhersage von Fällen mit positivem Ausgang und insgesamt bessere Ergebnisse als das logistische Einzelvariablen-Modell. Zudem bot die Search Partition Analysis nach dem hier vorliegenden erstmaligen Vergleich mit dem ebenfalls verteilungs- und annahmefreien Grobmengen-Modell neben einer besseren Klassifikationsleistung zudem eine höhere Akzeptanz des Modells für den Einsatz in der klinischen Praxis. Da dieses auf nominale Daten spezialisierte Modell außerdem auf die gegen hierarchische Partitionierungsverfahren oftmals angeführte Aufsplittung der Stichprobe verzichtet und in jedem Berechnungsschritt auf die ge-

samte Stichprobe zurückgreift und darüber hinaus die Randverteilung der abhängigen Variable zu berücksichtigen vermag, erscheint es für den Einsatz gerade bei klinischen Studien mit beschränkter Stichprobengröße und einem eher selten auftretenden Zielkriterium - wie etwa der Schizophrenie - gut geeignet.

Sowohl die logistische Regressions- als auch die Search Partition Analyse tendieren jedoch dazu, ausschließlich auf Variablen zurückzugreifen, die einen signifikanten Häufigkeitsunterschied für die abhängige Variable aufweisen, und unterschätzen damit möglicherweise die Rolle seltener, aber hoch prädiktiver Symptome bei der Vorhersage. Diese finden Berücksichtigung bei der Selektion über die positive diagnostische Likelihood Ratio, deren Selektions-cut-off-Modell neben dem logistischen Cluster-Modell zu der höchsten Gesamtklassifizierungsleistung in der Validierungsstichprobe führte und wie dieses und das Partitionsmodell keine signifikante Überanpassung an die Daten der Berechnungsstichprobe zeigte. Da die diagnostischen Likelihood Ratios, ebenso wie Sensitivität und Spezifität, unabhängig von der Prävalenz des Zielkriteriums in der Stichprobe sind, im Gegensatz zu den diagnostischen Effizienzindizes aber keine absolute Erkrankungswahrscheinlichkeit, sondern die Erhöhung bzw. Verringerung der Apriori-Erkrankungswahrscheinlichkeit anzeigen, stehen sie inhaltlich der klinischen Praxis, insbesondere der Prognostik näher und erlauben auch die Berücksichtigung nicht in das Modell aufgenommener Risikofaktoren. Somit bieten sie allgemein eine bessere Grundlage für die vergleichende Bewertung ermittelter Vorhersagemodelle und können möglicherweise verlässlichere Hinweise auf 'wichtige' Variablen als die diagnostischen Effizienzindizes geben. Ob und inwieweit eine Vorauswahl der in Analysen eingehenden unabhängigen Variablen nach ihrer diagnostischen Likelihood Ratio anstelle von signifikanten Häufigkeitsvergleichen zu besseren Vorhersageleistungen führt, werden allerdings zukünftige Studien zeigen müssen.

Es wird ebenfalls weiterer umfangreicher Studien bedürfen, ehe die hier und in anderen bislang vorliegenden Studien gefundenen Unterschiede in der Klassifikationsleistung der unterschiedlichen Verfahren als gesichert angesehen und verlässliche Anwendungsempfehlungen gegeben werden können. Dabei wäre es schon ein entscheidender Schritt nach vorn, wenn sich die Unterschiede der Klassifizierungsleistung der einzelnen Vorhersagemodelle in der Validierungsstichprobe auch an einer weiteren, neuen und völlig unabhängigen Stichprobe bestätigen ließen.

6 ZUSAMMENFASSUNG

In der Medizin findet zunehmend ein Paradigmenwandel von einem kurativen hin zu einem präventiven Ansatz statt. Dieser generelle Trend beeinflusst in den letzten Jahren verstärkt auch die psychiatrische Forschung, insbesondere auf dem Gebiet der schizophrenen Erkrankungen, für die eine oft mehrjährige Prodromalphase vor der Erstmanifestation schon seit langem bekannt ist. Doch wurde die prodromale Symptomatik einer schizophrenen Störung über Jahrzehnte generell als zu unspezifisch für die Diagnose und die damit verbundene Einleitung präventiver Behandlungsmaßnahmen angesehen, wodurch Bemühungen um eine Früherkennung nur sehr vereinzelt erfolgten. Die Früherkennungsforschung in diesem Gebiet wurde gleichwohl nicht nur durch das sehr vielfältige psychopathologische Erscheinungsbild eines initialen schizophrenen Prodroms erschwert, sondern es standen darüber hinaus auch lange Zeit kaum adäquate Analysemethoden für die in der psychopathologischen Forschung vorherrschenden Nominaldaten, mit Ausnahme einfacher Häufigkeitsvergleiche mittels χ^2 -Tests, zur Verfügung. Während jedoch bei der Entwicklung von nonparametrischen Verfahren zur Bestimmung von Vorhersagemodellen für eine dichotome abhängige Variable in den letzten zwei Jahrzehnten große Fortschritte erzielt wurden und eine ganze Reihe unterschiedlicher Methoden dem Wissenschaftler heute in gut handhabbarer Form zur Auswahl stehen, haben sich bisher nur sehr wenige Studien dem Vergleich der Klassifikations- bzw. Vorhersageleistung dieser Methoden gewidmet.

In der vorliegenden Studie wurden deshalb acht Ansätze zur Generierung von Vorhersagemodellen bei gleichzeitiger Variablenselektion einander hinsichtlich ihrer Klassifizierungsleistung in einer Validierungsstichprobe sowie ihrer klinischen Akzeptanz gegenübergestellt:

- eine kombinierte schrittweise und nonparametrische Diskriminanzanalyse von in χ^2 -Tests signifikant gewordenen unabhängigen Variablen,
- eine kombinierte schrittweise und nonparametrische Diskriminanzanalyse zuvor über die unabhängigen Variablen generierter Cluster,
- eine schrittweise logistische Regressionsanalyse von in χ^2 -Tests signifikant gewordenen unabhängigen Variablen,
- eine schrittweise logistische Regressionsanalyse zuvor über die unabhängigen Variablen generierter Cluster,
- eine auf dem Grob-Mengen-Modell basierte Analyse,
- ein nicht-hierarchisches Partitionierungsverfahren, die Search Partition Analysis,
- ein auf diagnostischen Effizienzindizes basierender Selektionsansatz mit Einführung eines cut-off-Wertes und
- ein auf positiven diagnostischen Likelihood Ratios basierender Selektionsansatz mit Einführung eines cut-off-Wertes.

Als Analysegrundlage, randomisiert aufgeteilt in eine 120 Patienten umfassende Berechnungs- und eine 40 Patienten umfassende Validierungsstichprobe, dienten die Daten der weltweit ersten, abgeschlossenen prospektiven Studie zur Früherkennung von Schizophrenie anhand 72 feiner, selbst wahrgenommener psychopathologischer Auffälligkeiten, wie sie detailliert und umfassend mit der 'Bonner Skala für die Beurteilung von Basissymptomen - BSABS' erhoben werden können. Aufgrund der bereits bestehenden Hinweise auf einen direkten Zusammenhang solcher Basissymptome und der Entwicklung einer schizophrenen Psychose, der nahezu gleichmäßigen Randverteilung der abhängigen Variable 'Übergang in eine Schizophrenie', der Abwesenheit fehlender Werte und nicht zuletzt des der Zielsetzung der verwandten Methoden angemessenen, prospektiven Designs wurde diese Stichprobe als gut geeignet für die vorliegende Fragestellung angesehen.

Die Ergebnisse bestätigten weitgehend die bisherigen Befunde methodischer Vergleichsstudien, die ebenfalls nur marginale Unterschiede in der Diskriminationsfähigkeit bzw. Klassifizierungsleistung verschiedener Verfahren fanden: Die kombinierte Diskriminanzanalyse der Einzelitems zeigte sich infolge der in der schrittweisen Analyse vorausgesetzten, aber fehlenden Normalverteilung allen anderen Verfahren in ihrer Klassifizierungsleistung mit 67,5% korrekter Gruppenzuordnungen in der Validierungsstichprobe unterlegen, die der Cluster jedoch zumindest der logistischen Regressionsanalyse der Einzelitems, dem Grob-Mengen-Modell und dem Selektionsmodell der diagnostischen Effizienzindizes ebenbürtig; die Variablenselektion über von der Prävalenz des Zielkriteriums in der Stichprobe unabhängige diagnostische Likelihood Ratios führte zu besseren und verlässlicheren Klassifizierungsleistungen als die über diagnostische Effizienzindizes; und das nicht-hierarchische Partitionsmodell war dem logistischen Modell der Einzelvariablen, nicht aber dem der Cluster überlegen, das gemeinsam mit dem Selektionsmodell über die positive diagnostische Likelihood Ratio mit 77,5% zutreffender Vorhersagen für die Validierungsstichprobe zu der besten Klassifizierungsleistung führte und wie diese, das Partitionsmodell und das diskriminanzanalytische Cluster-Modell keine signifikante Überanpassung an die Daten der Berechnungsstichprobe zeigte. Die Selektionsmodelle, das Partitionierungsmodell und das logistische Cluster-Modell fanden zudem auch noch eine zumindest zufriedenstellende klinische Akzeptanz bei den befragten Experten auf dem Gebiet der Früherkennung schizophrener Psychosen, wobei insbesondere die Selektionsmodelle sowohl hinsichtlich ihrer klinischen Plausibilität als auch ihrer Anwendbarkeit besonders positiv beurteilt wurden.

Darüber hinaus zeigte ein Vergleich der Vorhersageleistung der generierten Modelle für die einzelnen Patienten der Validierungsstichprobe, daß nur 40% von allen Modellen richtig und ganze 15% von allen falsch klassifiziert wurden, so daß nicht nur für 10%, sondern für 45% der Patienten die korrekte Vorhersage von dem applizierten Prognosemodell abhängig war.

Hierbei wurden auch von den beiden am besten klassifizierenden Verfahren, dem logistischen Clustermodell und dem Selektionsmodell der diagnostischen Likelihood Ratios, einzelne Patienten ausschließlich von ihnen fehlklassifiziert.

Im Hinblick auf die Variablenselektion fügten sich alle acht Modelle inhaltlich gut in die bestehenden Befunde der Basissymptom-Forschung und in heutige und frühere Entwicklungsmodelle schizophrener Episoden ein, indem sie vor allem kognitive Denkstörungen, aber auch Wahrnehmungsstörungen einbezogen. Für die insgesamt berücksichtigten 23 Symptome ergab sich hierbei ein Kernsatz von sieben kognitiven Denkstörungen, der zumindest in der Hälfte der Modelle aufgenommen wurde. Allein das mit 17 Variablen umfangreichste Selektionsmodell über die diagnostischen Likelihood Ratios enthielt dabei Variablen, die keinen signifikanten Häufigkeitsunterschied in der abhängigen Variable geboten hatten, und erscheint gleichsam als einziges Modell in der Lage, auch seltene, aber möglicherweise hoch prädiktive und für eine korrekte Prognose sehr aussagekräftige Symptome zu berücksichtigen.

Obwohl somit die vorliegende Studie die in der Literatur beschriebenen Unterschiede der Vorhersageleistung der untersuchten Methoden weitestgehend bestätigte, wird es noch weiterer Untersuchungen bedürfen, bis diese Trends als gesichert angesehen und detaillierte Anwendungsempfehlungen für die Forschungspraxis formuliert werden können. Auch wird abzuwarten bleiben, ob und inwieweit sich die generierten Modelle an unabhängigen, neu erhobenen Stichproben bewähren.

Bis dahin sollte den Anwendungsvoraussetzungen einzelner Modelle ein hohes Maß an Beachtung geschenkt werden, um dem Entstehen verzerrter oder gar fälschlicher Vorhersagemodelle vorzubeugen. Dabei sollte, wenn Unsicherheit über zugrunde liegende Modellannahmen besteht, auf annahmefreie Verfahren, wie etwa Partitionierungsansätze zurückgegriffen werden. Eine Berechnung dieser Modelle empfiehlt sich jedoch auch in Kombination mit logistischen Regressionsanalysen, da sie Hinweise auf mögliche wichtige Interaktionen von unabhängigen Variablen geben können, die von logistischen Regressionsverfahren nicht selbständig aufgedeckt werden können.

7 LITERATUR

- Addington D, Addington J, Patten S (1998). Depression in people with first-episode schizophrenia. *Br J Psychiatr* 172 (suppl. 33): 90-92.
- Addington J, Addington D (1998). Effect of substance misuse in early psychosis. *Br J Psychiatry* 172 (suppl. 33): 134-136.
- Agresti A (1996). *An Introduction to Categorical Data Analysis*. New York Chichester Brisbane Toronto Singapore, Wiley.
- American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders, third edition: DSM-III*. Washington, APA.
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders, third edition revised: DSM-III-R*. Washington, APA.
- American Psychiatric Association (1992). *Report from the DSM-IV field trial for schizophrenia and related psychotic disorders*. Iowa City, University of Iowa.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders, fourth edition: DSM-IV*. Washington, APA.
- Amin S, Brewin J, Singh SP, Harrison GL, Medley I, Kwiecinski R (1998). Diagnostic Stability of First Onset Psychosis: Implications for Epidemiology. Posterpräsentation auf dem Ninth Biennial Winter Workshop on Schizophrenia, Davos, Schweiz, 07.02. - 13.02.1998.
- Andreasen NC (1982). Negative symptoms in schizophrenia. *Arch Gen Psychiatr* 39: 784-788.
- Andreasen NC, Flaum M (1991). Schizophrenia: the characteristic symptoms. *Schizophr Bull* 17: 27-49.
- Andreß H-J, Hagenars JA, Kühnel S (1997). *Analyse von Tabellen und kategorialen Daten*. Berlin Heidelberg, Springer.
- Appleton DR (1995). What do we mean by a statistical model? *Stat Med* 14: 185-187.
- Arboleda C, Holzman PS (1985) Thought disorder in children at risk for psychosis. *Arch Gen Psychiatr* 42: 1004-1013.
- Armitage P (1971). *Statistical methods in medical research*. Oxford, Blackwell.
- Armitage P, Gehan EA (1974). Statistical methods of the identification and use of prognostic factors. *Int J Cancer* 13: 16-36.
- Ashby M, Neuhaus JM, Hauck WW, Bacchetti P, Heilborn DC, Jewell NP, Segal MR, Fusaro RE (1992). An annotated bibliography of methods for analysing correlated categorical data. *Stat Med* 11: 67-99.

- Backhaus K, Erichson B, Plinke W, Schuchard-Fischer, Ch, Weiber R (1987). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 4., neu bearb. u. erw. Aufl. Berlin Heidelberg, Springer.
- Barr CE, Mednick SA, Munk-Jørgenson JP (1990). Exposure to influenza epidemics during gestation and adult schizophrenia: A 40-year study. *Arch Gen Psychiatr* 47: 869-874.
- Bechdorf A, Halve S, Schultze-Lutter F, Klosterkötter J (1998). Selbst wahrnehmbare Vulnerabilität, Prodromalsymptome und Bewältigungsreaktionen vor schizophrenen und affektiven Episoden. *Fortschr Neurol Psychiatr* 66: 378-386.
- Begg CB, Greenes RA (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 39: 207-215.
- Beiser M, Erickson D, Fleming JAE, Iacono WG (1993). Establishing the Onset of Psychotic Illness. *Compr Psychiatr* 150: 1349-1354.
- Bell RQ (1992). Multiple-risk cohorts and segmenting risk as solutions to the problem of false positives in risk for the major psychoses. *Psychiatry* 55: 370-381.
- Berg S (1981). *Optimalität bei Clusteranalysen*. Dissertation an der Universität Münster.
- Berner P (1982). Achsensyndrome endogener Psychosen. In: Huber G (Hrsg.). *Endogene Psychosen: Diagnostik, Basissymptome und biologische Parameter*. Stuttgart, Schattauer: S. 77-91.
- Biggs D, De Ville B, Suen E (1991). A method of choosing multiway partitions for classification and decision trees. *J Appl Stat* 18: 49-64.
- Birchwood M (1992). Early intervention in schizophrenia: theoretical background and clinical strategies. *Br J Clin Psychol* 31: 257-278.
- Birchwood M, Macmillan JF (1993). Early intervention in schizophrenia. *Aust NZ J Psychiatr* 27: 374-378.
- Birchwood M, Smith J, Macmillan JF, Hogg B, Prasad R, Harvey C, Bering S (1989). Predicting relapse in schizophrenia: The development and implementation of an early signs monitoring system using patients and families as observers: A preliminary investigation. *Psychol Med* 19: 649-656.
- Bleuler E (1908). Die Prognose der Dementia praecox (Schizophreniegruppe). *Allg Ztschft Psychiatr Psychisch-Gerichtl Med* 65: 436-464.
- Bleuler E (1911). Dementia praecox oder Gruppe der Schizophrenie. In: Aschaffenburg G (Hrsg.). *Handbuch der Psychiatrie*, 4. Abtlg, 1. Hälfte. Leipzig, Deuticke.
- Bortz J (1985). *Lehrbuch der Statistik für Sozialwissenschaftler*. 2. vollst. neu bearb. u. erw. Aufl.. Berlin Heidelberg New York Tokyo, Springer.
- Boyko EJ (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test: short cut or wrong turn? *Med Decis Making* 14: 175-179.

- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984). *Classification and Regression Trees*. Belmont, CA, Wadsworth International.
- Brown S, Birtwistle J (1998). People with schizophrenia and their families. *Br J Psychiatr* 173: 139-144.
- Cameron DE (1938). Early schizophrenia. *Am J Psychiatr* 95: 567-578.
- Cannon TD, Mednick SA, Parnas J, Schulsinger F, Praestholm J, Verstergaard AA (1993). Developmental brain abnormalities in the offspring of schizophrenic mothers: I. Contributions of genetic and perinatal factors. *Arch Gen Psychiatr* 50: 551-564.
- Canstatt KFC (1841). *Handbuch der medicinischen Klinik*. 1.Bd. Die spezielle Pathologie und Therapie vom klinischen Standpunkt aus bearbeitet. Erlangen, Enke.
- Carpenter WT, Heinrichs DW (1983). Early intervention, limited, targeted pharmacotherapy of schizophrenia. *Schizophr Bull* 9: 533-542.
- Carpenter WT jun., Buchanan RW (1994). Schizophrenia. *N Engl J Med* 330: 681-690.
- Chapman J (1966). The early symptoms of schizophrenia. *Br J Psychiatr* 112: 225-251.
- Chapman LJ, Chapman JP (1987) The search for symptoms predictive of schizophrenia. *Schizophr Bull* 13: 497-504.
- Chapman LJ, Chapman JP, Raulin ML (1976). Scales for physical and social anhedonia. *J Abn Psychol* 85: 374-382.
- Chapman LJ, Chapman JP, Raulin ML (1978). Body-image aberration in schizophrenia. *J Abn Psychol* 87: 399-407.
- Christensen R (1981). *Entropy Minimax Sourcebook*. Volume 1: General Description. Lincoln MA, Entropy Limited.
- Clark LA, Pregibon D (1992). Tree-based models. In: Chambers JM, Hastie TJ (eds.). *Statistical models in S*. California, Wadsworth & Brooks: pp. 377-420.
- Cohen J (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, Academic Press.
- Conrad K (1958). *Die beginnende Schizophrenie*. Stuttgart, Thieme.
- Cook EF, Goldman L (1984). Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *J Chron Dis* 37: 721-731.
- Copas JB (1997). Using regression models for prediction: shrinkage and regression to the mean. *Stat Meth Med Res* 6: 167-183.
- Cornblatt B, Obuchowski M, Schnur D, O'Brien JD (1998). Hillside study of risk and early detection in schizophrenia. *Br J Psychiatr* 172 (suppl. 33): 26-32.
- Corpuz M, Hilton E, Lardis P, Singer C, Zolan J (1991). Problems in the use of serologic tests for the diagnosis of Lyme disease. *Arch Intern Med* 151: 1837-1840.

- Cox DR (1958). The regression analysis of binary sequences. *J R Stat Soc Ser B* 20: 215-242.
- Cox DR (1972). Regression models and life-tables (with discussion). *J Royal Stat Soc B* 34: 187-220.
- Cuesta MJ, Peralta V, Juan JA (1996) Abnormal subjective experiences in schizophrenia: its relationship with neuropsychological disturbances and frontal signs. *Eur Arch Psychiatr Clin Neurosci* 246: 101-105.
- Cutting J (1985). *The Psychology of Schizophrenia*. London, England, Churchill Livingstone.
- Cutting J, Dunne F (1989). Subjective experience of schizophrenia. *Schizophr Bull* 15: 217-231.
- DeBoeck P, Rosenberg S (1988). Hierarchical classes: Model and data analysis. *Psychometrika* 53: 361-381.
- Docherty JP, van Kammen DP, Siris SG, Marder SR (1978). Stages of the Onset of Schizophrenic Psychosis. *Am J Psychiatr* 135: 420-426.
- Domann T, Klug J (1979). Das Present-State-Examination (PSE) zur psychiatrischen Befunderhebung. In: Deutsche Gesellschaft für Verhaltenstherapie e. V. (Hrsg.). *Klinische Psychologie - Fortschritte in Diagnostik und Therapie - Kongreßbericht II - Hamburg 1978*. Tübingen, Mitteilungen der DGVT: S. 232-244.
- Doyle P (1973). The use of automatic interaction detector and similar search procedures. *Operational Res Quart* 24: 465-467.
- Ebel H, Gross G, Klosterkötter J, Huber G (1989). Basic symptoms in schizophrenic and affective psychoses. *Psychopathology* 22: 224-232.
- Eckblad M, Chapman LJ (1983). Magical ideation as an indicator of schizotypy. *J Consult Clin Psychol* 51: 215-225.
- Edell WS (1995). The Psychometric Measurement of Schizotypy Using the Wisconsin Scales of Psychosis Proneness. In: Miller GA (ed.). *The behavioral high-risk in psychopathology*. New York Berlin Heidelberg, Springer: 3-46.
- Einhorn HJ (1972). Alchemy in behavioural sciences. *Public Opin Quart* 36: 367-378.
- Eisenstein EL, Alemi F (1996). A Comparison of Three Techniques For Rapid Model Development: An Application in Patient Risk-Stratification. In: Cimino JJ (ed.). *Proceedings of the 1996 AMIA Annual Fall Symposium. Beyond the Superhighway: Exploiting the Internet with Medical Informatics*. Philadelphia, Hanley & Belfus: 443-447.
- Erlenmeyer-Kimling L, Cornblatt B (1987). High-Risk Research in Schizophrenia: A Summary of What Has Been Learned. *J Psychiatr Res* 22: 401-411.

- Falloon IRH (1992). Early intervention for first episodes of schizophrenia: A preliminary exploration. *Psychiatry* 55: 4-15.
- Feinstein R (1977). *Clinical Biostatistics*. Saint Louis, C.V. Mosby Co.: pp. 414-443.
- Fisher RA (1958). *Statistical methods for research workers*, 13th edn.. London, Oliver & Boyd.
- Flegg HG (1965). *Boolean Algebra and its Application*. Edinburgh London, Blackie: chapter 6.
- Fleiss JL, Spitzer RL, Cohen J, Endicott J (1977). Three computer diagnosis methods compared. *Arch Gen Psychiatr* 27: 643-649.
- Forsyth RS, Clarke DD & Wright RL (1944). Overfitting revisited: an information-theoretic approach to simplifying discrimination decision trees. *J Exp Theor Artificial Intell* 6: 289-302.
- Friedman JH, Roosen CR (1995). An introduction to multivariate adaptive regression splines. *Stat Meth Med Res* 4: 197-217.
- Genduso LA, Haley JC (1997). Cost of illness studies for schizophrenia: components, benefits, results, and implications. *Am J Managed Care* 3: 873-877.
- Goldman L, Weinberg M, Weisberg M, Olshen R, Cook EF, Sargent RK, Lamas GA, Dennis C, Wilson C, Deckelbaum L, Fineberg H, Stiratelli R (1982). A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 307: 586-596.
- Gottesman II (1991). *Schizophrenia Genesis. The origins of madness*. New York, Freeman.
- Gross G, Huber G (1989). Das Basissymptomkonzept idiopathischer Psychosen. *Zbl Neuro* 252: 655-673.
- Gross G, Huber G, Klosterkötter J, Linz M (1987). *Bonner Skala für die Beurteilung von Basissymptomen (BSABS; Bonn Scale for the Assessment of Basic Symptoms)*. Berlin Heidelberg New York, Springer.
- Gross G, Huber G, Klosterkötter J, Linz M (1992a) di Maggini C, Dalle Luche R (eds.) *Scala di Bonn per la valutazione dei sintomi di base*. Pisa, ETS Editrice.
- Gross G, Huber G, Klosterkötter J, Linz M (1992b) Moringio Escalante JC (ed.) *Escala de Bonn para Evaluación de Sintomas Básicos*. Asunción, Universidad Nacional de Asunción.
- Gross G, Huber G, Klosterkötter J, Linz M (1995) Parnas J, Handest P (eds.) *Bonnskalaen til vurdering af basissymptomer*. Kopenhagen, Mogens Handest.
- Gross G, Stassen HH, Huber G, Klosterkötter J (1990) Reliability of the psychopathological documentation scheme BSABS. In Stefanis CN, Rabavilas AD, Soldatos CR (eds.) *Psychiatry: A world perspective - Volume 1*. Amsterdam New York Oxford, Excerpta Medica: pp 199-203.
- Gruszecki G, Slowinski J, Stefanowski J (1990). *Rough DAS - Rough Sets-based Data Analysis System. User's Manual*. Warszawa, APRO SA.

- Grzymala-Busse J (1992). LERS - a system for learning from examples based on rough sets. In: Slowinski R (ed.). Intelligent decision support, Handbook of Application and Advances of the Rough Set Theory, vol. 3. Dordrecht, Kluwer: p. 3.
- Hadorn DC, Draper D, Rogers WH, Keeler EB, Brook RH (1992). Cross-validation performance of mortality prediction models. *Stat Med* 11: 475-489.
- Häfner H (1993). What is schizophrenia? *Neurol Psychiatr Brain Res* 2: 36-52.
- Häfner H, Maurer K (1996). Methodische Probleme und Lösungswege bei der Erfassung von Beginn und Frühverlauf psychischer Krankheiten am Beispiel Schizophrenie. In: Saß H (Hrsg.). *Psychopathologische Methoden und psychiatrische Forschung*. Jena Stuttgart, Fischer: S. 77-105.
- Häfner H, Nowotny B, Löffler W, an der Heiden W, Maurer K (1995). When and how does schizophrenia produce social deficits? *Eur Arch Psychiatr Clin Neurosci* 246: 17-28.
- Häfner H, Riecher A, Maurer K, Meissner S, Schmidtke A, Fätkenheuer B, Löffler W, an der Heiden W (1990). Ein Instrument zur retrospektiven Einschätzung des Erkrankungsbeginns bei Schizophrenie (Instrument for the retrospective assessment of the onset of schizophrenia - "IRAOS") - Entwicklung und erste Ergebnisse. *Zt Klin Psychol* 19: 230-255.
- Hager W (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In: Lürer G (Hrsg.). *Allgemeine Experimentelle Psychologie*. Stuttgart, Fischer: S. 43-264.
- Halperin M, Blackwelder WC, Verter JI (1971). Estimation of the multivariate logistic risk function: a comparison of the discriminant function and the maximum likelihood approaches. *J Chron Dis* 24: 125-158.
- Hambrecht M, Häfner H (1996). Substance abuse and the onset of schizophrenia. *Biol Psychiatr* 39: 1-9.
- Hanley JA, McNeill BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 50: 23-36.
- Harrell FE Jr, Lee KL, Matchar DB, Reichert TA (1985). Regression Models for Prognostic Prediction: Advantages, Problems, and Suggested Solutions. *Cancer Treat Rep* 69: 1071-1077.
- Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA (1982). Evaluating the yield of medical tests. *J Am Med Assoc* 247: 2543-2546.
- Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA (1984). Regression modelling strategies for improved prognostic prediction. *Stat Med* 3: 143-152.
- Harris EK, Albert A (1991). *Survivorship Analysis for Clinical Studies*. New York, Marcel Dekker.

- Hashemi R (1989). An automated rule generator (R^2) based on the rough sets. In: Gordon J (ed.). The 3rd Oklahoma Symposium on Artificial Intelligence. Tulsa OK, 1989: pp. 89-112.
- Hashemi R, Jelovsek FR, Razzaghi M (1993). Developmental Toxicity Risk Assessment: A Rough Sets Approach. *Meth Inform Med* 32: 47-54.
- Heinimaa MLA (2000). Some conceptual problems of early Intervention approach in schizophrenia. *Hongkong J Psychiatr* (in press).
- Heinrichs DW, Carpenter WT jun. (1985). Prospektive study of prodromal symptoms in schizophrenic relapse. *Am J Psychiatr* 142: 371-373.
- Heinrichs DW, Cohen BP, Carpenter WT jun. (1985). Early insight and the management of schizophrenic decompensation. *J Nervous Ment Dis* 173: 133-137.
- Helgason L (1990). Twenty years' follow-up of first psychiatric presentation for schizophrenia: What could have been prevented? *Acta Psychiatr Scand* 81: 231-235.
- Herz MI, Lamberti JS (1995). Prodromal symptoms and relapse prevention in schizophrenia. *Schizophr Bull* 21: 541-551.
- Herz MI, Melville C (1980). Relapse in schizophrenia. *Am J Psychiatr* 137: 801-805.
- Herz MI, Glazer W, Mirza M, Mostert M, Hafez H (1989). Treating prodromal episodes to prevent relapse in schizophrenia. *Br J Psychiatr* 155 (suppl. 5): 123-127.
- Hoffman RE, McGlashan TH (1993). Parallel distributed processing and the emergence of schizophrenic symptoms. *Schizophr Bull* 19: 119-140.
- Hoffman RE, Rapaport J, Ameli R, McGlashan TH, Harcherik D, Servan-Schreiber D (1995). A neural network simulation of hallucinated "voices" and associated speech perception impairments in schizophrenic patients. *J Cognitive Neurosc* 7: 479-496.
- Hogarty GE, Anderson CM, Reiss DJ (1986). Family education, social skills training, and maintenance chemotherapy in the aftercare of schizophrenia. *Arch Gen Psychiatr* 43: 633-642.
- Hollister JM, Laing P, Mednick SA (1996). Rhesus incompatibility as a risk-factor for schizophrenia in male adults. *Arch Gen Psychiatr* 53: 19-24.
- Huber G (1966). Reine Defektsyndrome und Basisstadien endogener Psychosen. *Fortschr Neurol Psychiat* 34: 409-426.
- Huber G (1968). Verlaufsprobleme schizophrener Erkrankungen. *Schweizer Arch Neurol Neurochirurg Psychiat* 101: 346-368.
- Huber G (1983) Das Konzept substratnaher Basissymptome und seine Bedeutung für Theorie und Therapie schizophrener Erkrankungen. *Nervenarzt* 54: 23-32.
- Huber G (1986). Psychiatrische Aspekte des Basisstörungskonzepts. In: Süllwold L, Huber G (Hrsg.). *Schizophrene Basisstörungen*. Berlin Heidelberg New York, Springer: S. 39-143.

- Huber G (1992). Konzeption der Einheitspsychose aus der Sicht der Basisstörungslehre. In: Mundt Ch, Saß H (Hrsg.). Für und wider die Einheitspsychose. Stuttgart New York, Thieme: S. 61-72.
- Huber G (1995). Prodrome der Schizophrenie. Fortschr Neurol Psychiat 63: 131-138.
- Huber G (1997). The heterogeneous course of schizophrenia. Schizophr Res 23: 177-185.
- Huber G, Gross G (1989) The concept of basic symptoms in schizophrenic and schizoaffective psychoses. Recenti Prog Med 80: 646-652.
- Huber G, Gross G, Schüttler R (1979) Schizophrenie. Verlaufs- und sozialpsychiatrische Langzeituntersuchung an den 1945-1959 in Bonn hospitalisierten schizophrenen Kranken. Berlin, Springer.
- Huber G, Gross G, Schüttler R, Linz M (1980) Longitudinal studies of schizophrenic patients. Schizophr Bull 6 : 592-605.
- Humphreys MS, Johnstone EC, Macmillan JF, Taylor PJ (1992). Dangerous behaviour preceding first admission for schizophrenia (see comments). Br J Psychiatr 161: 501-505.
- Hüttenlocher PR (1979). Synaptic density in the human frontal cortex: Developmental changes and effects of ageing. Brain Res 163: 195-205.
- Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992). Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization Ten-Country Study. Psychol Med (monograph suppl 20). Cambridge, University Press.
- Jackson HJ, McGorry PD, Dudgeon P (1995). Prodromal symptoms of schizophrenia in first-episode psychosis. Prevalence and specificity. Compr Psychiatr 36: 241-250.
- Jackson HJ, McGorry PD, McKenzie D (1994). The reliability of DSM-III prodromal symptoms in first-episode psychotic patients. Acta Psychiatr Scand 90: 375-378.
- Jackson HJ, McGorry PD, Dakis J, Harrigan S, Henry L, Mihalopoulos C (1996). The inter-rater and test-retest reliabilities of prodromal symptoms in first-episode psychosis. Aust NZ J Psychiatr 30: 498-504.
- Jaeger J, Bitter I, Czobor P, Volavka J (1990). The Measurement of Subjective Experience in Schizophrenia: The Subjective Deficit Syndrome Scale. Compr Psychiatr 31: 216-226.
- Janzarik W (1959). Dynamische Grundkonstellationen in endogenen Psychosen. Berlin, Springer.
- Janzarik W (1968). Schizophrene Verläufe. Eine strukturdynamische Interpretation. Berlin Heidelberg New York, Springer.
- Janzarik W (1980). Der schizoaffective Zwischenbereich. Nervenarzt 51: 272-279.
- Jimeno Bulnes N, Jimeno Valdes A, Code Lopez V (1991). Algunos datos sobre el interes del Inventario Psicopatologico de Frankfurt en el estudio psicopatologico del sindrome

- delirante. En: Arán (ed.). XV Reunion de la Sociedad Española de Psiquiatría Biológica, Vol. I. Córdoba, 1-3 novembre 1990. Madrid, Lidio de Actas: pg. 51-82.
- Johannessen JO, Larsen TK (unpublished). Duration of Untreated Psychosis (DUP): an important target for intervention in schizophrenia? Jan Olav Johannessen, Rogaland Psychiatric Hospital, 4325 Dale i Sandnes, Stavanger, Norway.
- Johnston MH, Holzman PS (1979). Assessing Schizophrenic Thinking. A Clinical and Research Instrument for Measuring Thought Disorder. San Francisco Washington London, Jossey-Bass.
- Johnstone EC, Crow TJ, Johnson AL, Macmillan JF (1986). The Northwick Park Study of first episodes of schizophrenia. I. Presentation of illness and problems relating to admission. *Br J Psychiatr* 148: 115-120.
- Jones PB, Tarrant CJ (1998). Vorläufersyndrome funktioneller Psychosen. *ZNS J* 18: 4-15.
- Johnstone EC, Macmillan JF, Frith, CD, Benn DK & Crow TJ (1990). Further investigation of the predictors of outcome following first schizophrenic episodes (see comments). *Br J Psychiatr* 157: 182-189.
- Jöreskog KG, Sörbom D (1993). New features in LISREL 8. Chicago, Scientific Software.
- Kalbfleisch JD, Prentice RL (1980). The Statistical Analysis of Failure Time Data. New York, Wiley.
- Kass GV (1980). An exploratory technique for investigating large quantities of categorical data. *Appl Stat* 29: 119-127.
- Keith SJ, Matthews SM (1991). The diagnosis of schizophrenia: A review of onset and duration issues. *Schizophr Bull* 17: 51-67.
- Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D (1993) The Roscommon Family Study IV: Affective illness, anxiety disorder and alcoholism in relatives. *Arch Gen Psychiatr* 50: 952-960.
- Klosterkötter J (1992). Die Entwicklung der schizophrenen Symptome ersten Ranges. *Fundamenta Psychiatrica* 6: 81-94.
- Klosterkötter J (1994) Warning Signs and Symptoms in Schizophrenia. In Saugstad LF (ed.) *Psychiatry and Mental Health - the Legacy of Ørnulv Ødegård*. Oslo, LFSF: pp. 113-135
- Klosterkötter J (1996). Das Vulnerabilitätskonzept bei schizophrenen Erkrankungen. In: Möller H-J, Deister A (Hrsg.). *Vulnerabilität für affektive und schizophrene Erkrankungen*. Wien New York, Springer: S. 11-21.
- Klosterkötter J (1998). Von der Krankheitsbekämpfung zur Krankheitsverhütung - Ist ein solcher Paradigmenwandel auch für schizophrene Störungen möglich? *Fortschr Neurol Psychiatr* 66: 366-377.

- Klosterkötter J, Hellmich M, Schultze-Lutter F (2000). Ist die Diagnose schizophrener Störungen schon in der initialen Prodromalphase vor der psychotischen Erstmanifestation möglich? *Fortschr Neurol Psychiatr* 68 (Sonderheft 1): 13-21.
- Klosterkötter J, Ebel H, Schultze-Lutter F, Steinmeyer EM (1996). Diagnostic validity of basic symptoms. *Eur Arch Psychiatr Clin Neurosc* 246: 147-154.
- Klosterkötter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F (2001). Diagnosing schizophrenia in the initial prodromal phase. *Arch Gen Psychiatr* 58: 158-164.
- Klosterkötter J, Albers M, Steinmeyer EM, Hensen A, Saß H (1994). The diagnostic validity of positive, negative, and basic symptoms. *Neurol Psychiatr Brain Res* 2: 232-238.
- Klosterkötter J, Schultze-Lutter F, Gross G, Huber G, Steinmeyer EM (1997). Early self-experienced neuropsychological deficits and subsequent schizophrenic diseases: an 8-year average follow-up prospective study. *Acta Psychiatr Scand* 95: 396-404.
- Klosterkötter J, Gross G, Huber G, Wieneke A, Steinmeyer EM, Schultze-Lutter F (1997). Evaluation of the 'Bonn Scale for the Assessment of Basic Symptoms - BSABS' as an instrument for the assessment of schizophrenia proneness: A review of recent findings. *Neurol Psychiatr Brain Res* 5: 137-150.
- Knuiman MW, Vu HTV, Segal MR (1997). An empirical comparison of multivariable methods for estimating risk of death from coronary heart disease. *J Cardiovasc Risk* 4: 127-134.
- Koreen AR, Siris SG, Chakos M, Alvir J, Mayerhoff D, Lieberman J (1993). Depression in First-Episode Schizophrenia. *Am J Psychiatr* 150: 1643-1648.
- Koss N, Feinstein AR (1971). Computer-aided prognosis II. Development of a prognostic algorithm. *Arch Internal Med* 127: 448-459.
- Kraemer HC (1992). *Evaluating Medical Tests. Objective and quantitative guidelines.* Newbury Park London New Dehli, Sage
- Kraepelin E (1896). *Psychiatrie - Ein Lehrbuch für Studierende und Ärzte (5. Aufl.).* Leipzig, A. Barth.
- Kremen WS, Seidman LJ, Pepple JR, Lyons MJ, Tsuang MT, Faraone SV (1994). Neuropsychological Risk Indicators for Schizophrenia: A Review of Family Studies. *Schizophr Bull* 20: 103-119.
- Kryspin-Exner J, Lutterotti R (1982). Clusteranalytische Untersuchung über die Symptome des Frankfurter Beschwerdefragebogens. *Neuropsychiatr Clin* 1: 29-41.
- Kukla F, Gold R (1991). Schizophrene Minussymptomatik und ihre Erfassungsmethodik. *Fortschr neurol Psychiat* 59: 60-66.

- Larsen TK, Johannessen JO, Opjordsmoen S (1998). First Episode Schizophrenia with Long Duration of Untreated Psychosis. *Br J Psychiatr* 172 (suppl. 33): 45-52.
- Larsen TK, McGlashan TH, Moe LC (1996). First-episode schizophrenia: I. Early course parameters. *Schizophr Bull* 22: 241-256.
- LeBlanc M, Crowley J (1992). Relative risk trees for censored survival data. *Biometrics* 48: 411-425.
- Leff J (1987). A model of schizophrenic vulnerability to environmental factors. In: Häfner H, Gattaz WF, Janzarik W (eds.). *Search for the causes of schizophrenia*. Berlin Heidelberg New York Tokyo, Springer: 317-330.
- Leisenring W, Pepe MS (1998). Regression modelling of diagnostic Likelihood Ratios for the evaluation of medical diagnostic tests. *Biometrics* 54: 444-452.
- Lette J, Coletti BW, Cerino M, McNamara D, Eybalin M-C, Levasseur A, Nattel S (1994). Artificial Intelligence Versus Logistic Regression Statistical Modelling to Predict Cardiac Complications after Noncardiac Surgery. *Clin Cardiol* 17: 609-614.
- Lewis SE, Murray RM (1987). Obstetric complications, neurodevelopmental deviance, and risk of schizophrenia. *J Psychiatr Res* 21: 413-421.
- Lieberman JA, Kinon BJ, Loebel AD (1990). Dopaminergic mechanisms in idiopathic and drug-induced psychoses. *Schizophr Bull* 16: 97-110.
- Loebel AD, Lieberman JA, Alvir JM, Mayerhoff DI, Geisler SH, Szymanski SR (1992). Duration of psychosis and outcome in first-episode schizophrenia. *Am J Psychiatr* 149: 1183-1188.
- Loh W, Vanichsekatul N (1988). Tree-structured classification via generalised discriminant analysis. *J Am Stat Assoc* 83: 715-725.
- Long WJ, Griffith JL, Selker HP, D'Agostino RB (1993). A comparison of logistic regression to decision tree introduction in medical domain. *Comput Biomed Res* 26: 74-97.
- Mabbett A, Stone M, Washbrook J (1980). Cross-validatory section of binary variables in differential diagnosis. *Appl Stat* 29: 198-204.
- Malla AK, Norman RMG (1994). Prodromal Symptoms in Schizophrenia. *Br J Psychiatr* 164: 487-493.
- Malla AK, Norman RMG (1999). Facing the Challenges of Intervening Early in Psychosis. *Annals RCPSC* 32: 394-397.
- Malla AK, Norman RMG, Voruganti LP (1999). Improving outcome in schizophrenia: the case for early intervention. *CMAJ* 160: 843-846.
- Marder SR, Mintz J, Van Putten T, Lebell M, Wirshing WC, Johnston-Cronk K (1991). Early prediction of relapse in schizophrenia: an application of receiver operating characteristic (ROC) methods. *Psychopharmacol Bull* 27: 79-82.

- Marshall G, Grover FL, Henderson WG, Hammermeister KE (1994). Assessment of predictive models for binary outcomes: an empirical approach using operative death from cardiac surgery. *Stat Med* 13: 1501-1511.
- Marshall RJ (1985). An algorithm to generate a class of regular dichotomous partitions. Internal Report, Unit for Cancer Research, Leeds University.
- Marshall RJ (1986). Partitioning methods for classification and decision making in medicine. *Stat Med* 5: 517-526.
- Marshall RJ (1995). A program to implement a search method for identification of clinical subgroups. *Stat Med* 14: 2645-2659.
- Marshall RJ (1997). SPAN for Windows 95. A manual for SPAN. Department of Community Health, University of Auckland, New Zealand (<http://www.auckland.ac.nz/mch/span.htm>).
- Marshall RJ, Chisholm EM (1985). Hypothesis testing in the polychotomous logistic model with an application to detecting gastrointestinal cancer. *Stat Med* 4: 337-344.
- Marubini E, Morabito A, Valsecchi MG (1983). Prognostic factors and risk groups: some results given by using an algorithm suitable for censored survival data. *Stat Med* 2: 295-303.
- Maß R, Krausz M, Gross J (1995). Basissymptome bei Schizophrenie und Alkoholismus. *Nervenarzt* 66: 331-337.
- Maß R, Hirschfeld K, Wall E, Wagner H-B (1997). Validität der Erfassung schizophrener Basissymptome. *Nervenarzt* 68: 205-211.
- Maurer K, Häfner H (1995). Methodological aspects of the onset assessment in schizophrenia. *Schizophr Res* 15: 265-276.
- McGhie A, Chapman J (1961). Disorders of attention and perception in early schizophrenia. *Br J Med Psychol* 34: 103-115.
- McGlashan TH (1996). Early Detection and Intervention in Schizophrenia: Research. *Schizophr Bull* 22: 327-345.
- McGlashan TH, Johannessen JO (1996). Early Detection and Intervention with Schizophrenia: Rationale. *Schizophr Bull* 22: 201-222.
- McGorry PD (1994). The influence of illness duration on syndrome clarity and stability in functional psychosis: Does the diagnosis emerge and stabilize with time? *Aust NZ J Psychiatr* 28: 607-619.
- McGorry PD (1998). "A stitch in time"...the scope for preventive strategies in early psychosis. *Eur Arch Psychiatry Clin Neurosc* 248: 22-31.
- McGorry PD, Edwards J (1997). Early psychosis training pack. Victoria Mill / Australia, Gardiner-Caldwell Communications.
- McGorry PD, Copolov DL, Singh BS (1990). Royal Park Multidiagnostic Instrument for Psychosis: Part I. Rationale and Review. *Schizophr Bull* 16: 501-515.

- McGorry PD, Edwards J, Mihalopoulos SM (1996). EPPIC: An evolving system of early detection and optimal management. *Schizophr Bull* 22: 305-326.
- McGorry PD, Phillips L, Crump N, Yung A, Adlard S, McDonald A (1999). Predicting early transition to psychosis in young people at high risk. *Current Opinion Psychiatr* 12 (suppl. 1): 254.
- McGorry PD, Singh BS, Copolov DL, Kaplan I, Dossetor CR, van Riel RJ (1990). Royal Park Multidiagnostic Instrument for Psychosis: Part II. Development, Reliability, and Validity. *Schizophr Bull* 16: 516-536.
- McGorry PD, McFarlane C, Patton GC, Bell R, Hibbert ME, Jackson HJ, Bowes G (1995). The prevalence of prodromal features of schizophrenia in adolescence: a preliminary survey. *Acta Psychiatr Scand* 92: 241-249.
- McGorry PD, Phillips LJ, Yung AR, Francey S, Velakoulis D, Brewer W, Yuen HP, Hallgren M, Patton G, Adlard S, Hearn N, Crump N, Pantelis C (1999). The identification of predictors of psychosis in a high risk group. *Schizophr Res* 36 (Nos 1-3): 49-50.
- McNeil BJ, Keeler E, Adelstein SJ (1975). Primer on certain elements of medical decision making. *N Engl J Med* 293: 211-215.
- Mechelen I van, DeBoeck P (1990). Projection of a binary criterion into a model of hierarchical classes. *Psychometrika* 55: 677-694.
- Meehl PE (1962). Schizotaxia, schizotypy, schizophrenia. *Am Psychol* 17: 827-838.
- Meehl PE (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *J Pers Dis* 4: 1-99.
- Miller TJ, McGlashan TH, Woods SW, Stein K, Driesen N, Corcoran CM, Hoffman R, Davidson L (1999). Symptom assessment in schizophrenic prodromal states. *Psychiatric Quarterly* 70: 273-287.
- Milligan GW (1980). An Examination of the Effects of Six Types of Error Perturbation on Fifteen Clustering Algorithms. *Psychometrika* 50: 159-179.
- Milligan GW (1981). A Review of Monte Carlo Tests of Cluster Analysis. *Multivariate Beh Res* 16: 379-407.
- Moldin SO, Erlenmeyer-Kimling L (1994). Measuring Liability to Schizophrenia: Progress Report 1994: Editors Introduction. *Schizophr Bull* 20: 25-29.
- Morgan JN, Sonquist JA (1963). Problems in the analysis of survey data, and a proposal. *J Am Stat Assoc* 58: 415-434.
- Mundt Ch, Kasper S (1987). Zur Schizophreniespezifität von negativen und Basisstörungen. *Nervenarzt* 58: 489-495.
- Murray RM (1994). Neurodevelopmental Schizophrenia: The rediscovery of Dementia Praecox. *Br J Psychiatry* 165 (suppl. 25): 6-12.

- Newman LG, Waller J, Palestro CJ, Schwartz M, Klein MJ, Hermann G, Harrington E, Harrington M, Roman SH, Stangnaro-Green A (1991). Unsuspected osteomyelitis in diabetic foot ulcers. Diagnosis and monitoring by leukocyte scanning with indium in 111 oxyquinoline. *JAMA* 266: 1246-1251.
- Nuechterlein KH (1987). Vulnerability models for schizophrenia: state of the art. In: Häfner H, Gattaz WF, Janzarik W (eds.). *Search for the causes of schizophrenia*. Berlin Heidelberg New York Tokyo, Springer: 297-316.
- Nuechterlein KH, Dawson ME (1984). A heuristic vulnerability / stress model of schizophrenic episodes. *Schizophr Bull* 10: 300-312.
- Nuechterlein KH, Zaucha KM (1990). Similarities between information-processing abnormalities of actively symptomatic schizophrenic patients and high-risk children. In: Straube ER, Hahlweg K (eds.). *Schizophrenia. Concepts, vulnerability, and intervention*. Berlin Heidelberg New York Tokyo, Springer: 77-96.
- Nuechterlein KH, Dawson ME, Gitlin M, Ventura J, Goldstein MJ, Snyder KS, Yee CM, Mintz J (1992). *Developmental Processes in Schizophrenic Disorders: Longitudinal Studies of Vulnerability and Stress*. *Schizophr Bull* 18: 387-425.
- Øhrn A, Komorowski J (1997). ROSETTA - A Rough Set Toolkit for Analysis of Data. *Proc. Third International Joint Conference on Information Sciences, Durham, NC, USA, March 1-5, Vol. 3: pp. 403-407*.
- Øhrn A, Komorowski J, Skowron A, Synak P (1998). The Design and Implementation of a Knowledge Discovery Toolkit based on Rough Sets - The ROSETTA System. In: Polkowski L, Skowron A (eds.). *Rough Sets in Knowledge Discovery*. Physica.
- Olin SS, Mednick SA (1996). Risk Factors of Psychosis: Identifying Vulnerable Populations Premorbidly. *Schizophr Bull* 22: 223-240.
- Olin SS, Raine A, Cannon TD, Parnas J, Schulsinger F, Mednick SA (1997). Childhood Behavior Precursors of Schizotypal Personality Disorder. *Schizophr Bull* 23: 93-103.
- Parnas J (1999). From predisposition to psychosis: progression of symptoms in schizophrenia. *Acta Psychiatr Scand* 99 (suppl. 395): 20-29.
- Parnas J, Mednick SA (1991) Early predictors of onset and course of schizophrenia and schizophrenia spectrum. In Häfner H, Gattaz WF (eds.) *Search for the causes of schizophrenia*, Vol 2. Berlin Heidelberg New York, Springer: pp. 34-47.
- Pascual MR, Podriguez A, Zayas L, Moreno L, Lage A (1983). Factors associated with prognosis in human breast cancer. II. Multivariate stratification analysis. *Neoplasma* 30: 485-492.
- Pawlak Z (1982). Rough sets. *Int J Inform Comput Sci* 11: 145-177.

- Pawlak Z (1984). Rough classification. *Int J Man-Machine Studies* 20: 469-483.
- Pawlak Z (1986). Rough classification of patients after highly selective vagotomy for duodenal ulcer. *Int J Man-Machine Studies* 24: 413-433.
- Pawlak Z, Wong SKM, Ziarko W (1988). Rough sets: probabilistic versus deterministic approach. *Int J Man-Machine Studies* 29: 81-95.
- Preisser JS, Koch GG (1997). Categorical Data Analysis in Public Health. *Annu Rev Public Health* 18: 51-82.
- Quinlan JR (1983). Learning efficient classification procedures and their application to chess end games. In: Michalski RS, Carbonell JG, Mitchell TM (eds.). *Machine learning: An artificial intelligence approach*. Palo Alto, CA, Tioga Publishing: pp. 463-482.
- Quinlan JR (1993). C4.5. In: *Programs for Machine Learning*. San Mateo, CA, Morgan Kaufman.
- REDUCT System Inc. (1991-1992). *DataLogic / R and DataLogic / R+, Version 1.1*. REDUCT System Inc., Regina, Saskatchewan, Canada.
- REDUCT System Inc. (1991-1994). *DataLogic / R 1.3 and DataLogic / R+ 1.5*. REDUCT System Inc., Regina, Saskatchewan, Canada.
- Resch F (1992). *Therapie der Adoleszentenpsychosen. Psychopathologische, psychobiologische und entwicklungspsychologische Aspekte aus therapeutischer Sicht*. Stuttgart New York, Georg Thieme.
- Ricca V, Galassi F, La Malfa G, Mannucci E, Barciulli E, Carbras PL (1997). Assessment of basic symptoms in schizophrenia, schizoaffective and bipolar disorders. *Psychopathology* 30: 53-58.
- Rösler M, Bellaire W, Hengesch G, Kiesling-Muck H, Carls W (1985) Die uncharakteristischen Basissymptome des Frankfurter Beschwerdefragebogens und ihre Beziehung zu psychopathologischen Syndromen. *Nervenarzt* 56: 259-264.
- Sackett DI (1991). Clinical reality, binary models, babies and bath water. *J Clin Epidemiol* 44: 217-219.
- Sackett DI (1992). A primer on the precision and accuracy of the clinical examination. *JAMA*, 267: 2638-2644.
- Sackett DI, Haynes RB, Tugwell P (1985). *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, MA, Little & Brown.
- SAS Institute Inc. (1989). *SAS/STAT[®] User's Guide, Version 6, Fourth Edition, Volume 1*. SAS Institute Inc., Cary, NC, USA.

- Schahn J, Amelang M (1992). Mittelwertsunterschiede zwischen Selbst- und Fremdbeurteilungen: Eine vernachlässigte Größe? *Diagnostika* 18: 187-208.
- Schneider K (1946). *Klinische Psychopathologie*. Stuttgart, Thieme.
- Schneider K (1992). *Klinische Psychopathologie*, 14. Unveränderte Auflage mit einem Kommentar von Huber G und Gross G. Stuttgart New York, Thieme.
- Schouten HJA (1982). Measuring pairwise agreement among many observers. Some improvements and additions. *Biometric J* 24: 431-435.
- Schultze-Lutter F (2000). Ruling in or ruling out schizophrenic prodrome: what criteria for symptom selection should be used?. *Schizophr Res* 41: 179.
- Schultze-Lutter F, Klosterkötter J (1995). Do basic symptoms provide a possible explanation for the elevated risk for schizophrenia among mentally retarded? *Neurol Psychiatr Brain Res* 3: 29-34.
- Schultze-Lutter F, Steinmeyer EM, Klosterkötter J (2000). Early detection of schizophrenia. *Schizophr Res* (in press).
- Segal MR (1988). Regression trees for censored data. *Biometrics* 44: 35-47.
- Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB (1995). A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. *J Invest Med* 43: 468-476.
- Spitzer RL, Endicott J (1968). DIAGNO: A computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. *Arch Gen Psychiatr* 18: 746-756.
- Spitzer RL, Endicott J (1969). DIAGNO II: Further development in a computer program for psychiatric diagnosis. *Am J Psychiatr* 125 (suppl. 7): 12-21.
- Stirling J, Tantam D, Thomas P, Newby D, Montague L, Ring N, Rowe S (1991). Expressed emotion and early onset schizophrenia: a one year follow-up. *Psychol Med* 21: 675-685.
- Stirling J, Tantam D, Thomas P, Newby D, Montague L, Ring N, Rowe S (1993). Expressed emotion and schizophrenia: the ontogeny of EE during an 18-month follow-up. *Psychol Med* 23: 771-778.
- Strakowski SM, Keck PE, McElroy SL, Lonczak HS, West SA (1995). Chronology of Comorbid and Principal Syndromes in First-Episode Psychosis. *Compr Psychiatr* 36: 106-112.
- Süllwold L (1977). *Symptome schizophrener Erkrankungen. Uncharakteristische Basisstörungen*. Berlin Heidelberg New York, Springer.
- Süllwold L (1986). Die Selbstwahrnehmung defizitärer Störungen. *Psychologische Aspekte des Basisstörungskonzepts*. In Süllwold L, Huber G (Hrsg.). *Schizophrene Basisstörungen*. Berlin Heidelberg New York, Springer: S. 1-38.
- Süllwold L (1991). *Manual zum Farnkfurter Beschwerde-Fragebogen (FBF)*. Berlin Heidelberg New York, Springer.

- Süllwold L, Huber G (1986.). Schizophrene Basisstörungen. Berlin Heidelberg New York, Springer.
- Susser ES, Shang P (1992). Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944-1945. Arch Gen Psychiatr 49: 983-988.
- Teusch L (1984). Ein kritischer Beitrag zur Diagnostik der sog. Basisstörungen mit dem Frankfurter Beschwerdefragebogen (FBF). In: Hopf A, Beckmann A (Hrsg.). Forschungen zur biologischen Psychiatrie. Berlin Heidelberg New York, Springer: S. 309-315.
- Teusch L (1985) Substratnahe Basisstörungen oder nosologisch vieldeutige subjektive kognitive Störbarkeit? Nervenarzt 56: 265-269.
- Tienari P, Sorri A, Lahti I, Naarala M, Wahlberg KE, Ronkkot Pohjola J, Moring J (1985). The Finnish Adoption Family Study of Schizophrenia. Yale J Biol Med 58: 227-237.
- Truetti J, Cornfield J, Kannel W (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. J Chron Dis 20: 511-524.
- Tsumoto S, Tanaka H (1995). PRIMEROSE: Probabilistic Rule Induction Method based on Rough Sets and Resampling Methods. Comput Intelligence 11: 389-405.
- Turner TH (1989). Schizophrenia and mental handicap: an historical review, with implications for further research. Psychol Med 19: 301-314.
- Vaglum P (1996). Earlier Detection and Intervention in Schizophrenia: Unsolved Questions. Schizophr Bull 22: 347-351.
- van Os J (1998). Prevalence of hallucinations and delusions in the general population. Vortrag an der Klinik für Psychiatrie und Psychotherapie der Universität zu Köln, 12.08.1998.
- van Os J, Verdoux H, Maurice-Tison S, Gay B, Liraud F, Salamon R, Bourgois ML (1999). Self-reported psychosis-like symptoms and the continuum of psychosis. Soc Psychiatr Psychiat Epidem 34: 459-463.
- Verdoux H, Maurice-Tison S, Gay B, van Os J, Salamon R, Bourgois ML (1998). A survey of delusional ideation in primary-care patients. Psychol Med 28: 127-134.
- Weinstein MC, Feinberg HV, Elstein AS, Frazier HS, Neuhauser D, Neutra RR, McNeil BJ (1980). Clinical Decision Analysis. Philadelphia, W. B. Saunders.
- Weltgesundheitsorganisation, WHO (1973). Report of the International Pilot Study of Schizophrenia. Geneva, World Health Organization.
- World Health Organization, WHO (1975). Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death. Based on the Recommendations of the Ninth Revision Conference, 1975, and Adopted by the Twenty-ninth World Health Assembly. Geneva, World Health Organization.

- WHO (1993). Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F), klinisch-diagnostische Leitlinien, 2. Auflg.. Hrsg. v. Dilling H, Mombour W, Schmidt MH. Bern Göttingen Toronto Seattle, Huber.
- Williams R, Dickson RA (1995). Economics of schizophrenia. *Can J Psychiatr* 40 (suppl. 2): 60-67.
- Wing JK (1970). A standard form of psychiatric present state examination and a method for standardizing the classification of symptoms. In: Hare EH, Wing JK (eds.). *Psychiatric Epidemiology: An International Symposium*. London, Oxford University Press: pp. 93-108.
- Wing JK (1976). A technique for study psychiatric morbidity in in-patient and out-patient series and in general population samples. *Psychol Med* 6: 665-671.
- Wing JK, Cooper JE, Sartorius N (1974) Measurement and classification of psychiatric symptoms. An introduction manual for the PSE and Catego-program. London, Cambridge University Press.
- Wing JK, Cooper JE, Sartorius N (1982). Die Erfassung und Klassifikation psychiatrischer Symptome. Beschreibung und Glossar des PSE (Present State Examination) - ein Verfahren zur Erhebung des psychopathologischen Befundes. (Deutsche Bearbeitung: M. von Cranach). Weinheim, Beltz.
- Winter G (1997). Neuronale Netzwerkanalysen in der psychoiatrischen Forschung. BIOMED-1, Projekt ANDEE. Vortrag auf dem Treffen der AGNP-Arbeitsgemeinschaft 'Methoden', Mainz, 31.01.1997.
- Wyatt RJ (1991). Neuroleptics and the natural course of schizophrenia. *Schizophr Bull* 17: 325-351.
- Yap RH, Clarke DM (1996). An Expert System for Psychiatric Diagnosis using DSM-III-R, DSM-IV and ICD-10 Classifications. In: Cimino JJ (ed.). *Proceedings of the 1996 AMIA Annual Fall Symposium. Beyond the Superhighway: Exploiting the Internet with Medical Informatics*. Philadelphia, Hanley & Belfus: 229-233.
- Yung AR, McGorry PD (1996). The Prodromal Phase of First-Episode Psychosis: Past and Current Conceptualizations. *Schizophr Bull* 22: 353-370.
- Yung AR, McGorry PD (1997). Is pre-psychotic intervention realistic in schizophrenia and related disorders? *Aust NZ J Psychiatr* 31: 799-805.
- Yung AR, McGorry PD, McFarlane CA, Jackson HJ, Patton GC, Rakkar A (1996). Monitoring and Care of Young People at Incipient Risk of Psychosis. *Schizophr Bull* 22: 283-303.
- Yung AR, Phillips LJ, McGorry PD, McFarlane CA, Francey S, Harrigan S, Patton GC, Jackson HJ (1998). Prediction of psychosis. *Br J Psychiatr* 172 (suppl. 33): 14-20.

Ziarko W (1987). On reduction of knowledge presentation. Proceedings of the 2nd ACM SIGART International Symposium on Methodologies for Intelligent Systems. Charlotte 1987.

Zubin J (1990). Ursprünge der Vulnerabilitätstheorie. In: Olbrich R (Hrsg.). Therapie der Schizophrenie. Stuttgart Berlin Köln, Kohlhammer: 43-52.

ANHANG

ANHANG A: Bonner Skala für die Beurteilung von Basissymptomen - BSABS (Gross, Huber, Klosterkötter & Linz, 1987): Erhebungsbogen

BSABS-Item-Nr.	BASISSYMPTOM	?
A	DYNAMISCHE DEFIZIENZEN MIT DIREKTEN MINUS-SYMPTOMEN	
A.1	Erhöhte Erschöpfbarkeit	
A.1.1	Körperliche und seelisch-geistige Ermüdung	
A.1.2	Erschöpfbarkeit und Ermüdbarkeit mit Indirekten Minus-Symptomen, IMS	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
A.2	Erhöhtes Schlafbedürfnis	
A.3	Minderung an Spannkraft, Energie, Ausdauer, 'Geduld'	
A.3.1	Minderung an Spannkraft und Energie	
A.3.2	Minderung an Ausdauer und Geduld	
A.4	Minderung an Antrieb, Aktivität, Schwung, Elan, Initiative	
A.5	Mangelnde Entscheidungsfähigkeit, Entschlußschwäche, Unschlüssigkeit	
A.6	Affektive Veränderungen	
A.6.1	Veränderungen von Grundstimmung und emotionaler Resonanzfähigkeit	
A.6.2	Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten	
A.6.3	Abschwächung bejahender Fremdwert- und Sympathiegefühle	
A.6.4	Minderung des Kontaktbedürfnisses	
A.6.5	Phasenhafte depressive Verstimmungen	
	Hypomanische bzw. maniforme Phasen ?	
	Coenästhesien ?	
A.7	Störung der Kontaktfähigkeit und des In-Erscheinung-Tretens	
A.7.1	Minderung der Kontaktfähigkeit bei vorhandenem Kontaktwunsch	
A.7.2	Störung des In-Erscheinung-Tretens	
A.8	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten Stressoren	
A.8.1	... gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen	

BSABS-Item-Nr.	BASISSYMPTOM	?
A.8.2	... gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen	
	s1: Unterhaltung, Besuche	
	s2: Gegenwart zu vieler Menschen, "Trubel", "Rummel", Kaufhäuser	
	s3: Optische und/oder akustische Stimulation	
	s4: Andere Situationen?	
A.8.3	... gegenüber Arbeit unter Zeitdruck und/oder rasch wechselnden unterschiedlichen Anforderungen	
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	
B	DYNAMISCHE DEFIZIENZEN MIT INDIRECTEN MINUS-SYMPTOMEN	
B.1	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten Stressoren	
B.1.1	... gegenüber körperlicher und/oder psychischer arbeitsmäßiger Beanspruchung	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.1.2	... gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.1.3	... gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen	
	s1: Unterhaltung, Besuche	
	s2: Gegenwart zu vieler Menschen, "Trubel", "Rummel", Kaufhäuser	
	s3: Optische und/oder akustische Stimulation	
	s4: Andere Situationen?	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	

BSABS-Item-Nr.	BASISSYMPTOM	?
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.1.4	... gegenüber Arbeit unter Zeitdruck und/oder rasch wechselnden unterschiedlichen Anforderungen	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.1.5	...gegenüber Witterungseinflüssen	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.1.6	... gegenüber emotional affizierenden Ereignissen mit IMS obligat und gewöhnlich ausschließlich in Form von Coenästhesien	
B.2	Erhöhte Beeindruckbarkeit; erhöhte Erregbarkeit	
B.2.1	Erhöhte Beeindruckbarkeit durch alltägliche Ereignisse	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.2.2	Erhöhte Beeindruckbarkeit durch Verhaltensweisen anderer, die die Person persönlich betreffen	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	

BSABS-Item-Nr.	BASISSYMPTOM	?
B.2.3	Erhöhte Beeindruckbarkeit durch 'fremdes Leid'	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
B.2.4	Erhöhte Erregbarkeit und Reizbarkeit	
B.3	Erhöhte Reflexivität, Zwang, Phobie, autopsychische Depersonalisation	
B.3.1	Erhöhte Reflexivität: Verlust an Naivität, Unbekümmertheit, Unbefangenheit	
B.3.2	Zwangsphänomene	
B.3.3	Phobien	
B.3.4	Autopsychische Depersonalisation	
C	KOGNITIVE DENK-, WAHRNEHMUNGS- UND HANDLUNGSSTÖRUNGEN	
C.1	Kognitive Denkstörungen	
C.1.1	Gedankeninterferenz	
C.1.2	Zwangähnliches Perseverieren bestimmter Bewußtseinsinhalte	
	s1: Zurückliegende Vorgänge	
	s2: Zukunftsbezogene Bewußtseinsinhalte	
C.1.3	Gedankendrängen, Gedankenjagen	
C.1.4	Blockierung des jeweiligen Gedankenganges	
	s1: Reine Blockierung	
	s2: Reines Fading	
	s3: Gedankenschwund und -interferenz simultan	
	s4: Gedankenblockierung und -interferenz sukzessiv	
	s5: 'Fadenverlieren'	
C.1.5	Störung der Konzentrationsfähigkeit	
C.1.6	Störung der rezeptiven Sprache	
	s1: Visuell	
	s2: Akustisch	
C.1.7	Störung der expressiven Sprache	
C.1.8	Störung des unmittelbaren Behaltens	
C.1.9	Störung des Kurzzeitgedächtnisses	
C.1.10	Besonders strukturierte Störungen des Langzeitgedächtnisses	

BSABS-Item-Nr.	BASISSYMP TOM	?
C.1.11	Nicht (unter C.1.8 .- 10.) rubrizierbare Gedächtnisstörungen	
C.1.12	Verlangsamung und Erschwerung der Denkvorgänge	
C.1.13	Störung der Denkinitiative und gedankliche Intentionalität	
C.1.14	Störung der Revisualisation (der Vorstellung vertrauter Dinge)	
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	
	s1: Von Vorstellungen und Wahrnehmungen	
	s2: Von Phantasie- und Erinnerungsvorstellungen	
C.1.16	Störung der Symbolerfassung (Konkretismus)	
C.1.17	'Subjekt-Zentrismus' - Eigenbeziehungstendenz	
C.2	Kognitive Wahrnehmungsstörungen	
C.2.1	Verschwommen- und Trübsehen, passagere Blindheit, partielles Sehen	
	s1: Verschwommen-, Trübsehen	
	s2: Passagere Blindheit	
	s3: Partielles Sehen	
C.2.2	Lichtüberempfindlichkeit, Überempfindlichkeit gegenüber bestimmten visuellen Reizen; Photopsien	
	s1: Überempfindlichkeit gegen Licht oder visuelle Reize	
	s2: Photopsien	
C.2.3	Andere optische Wahrnehmungsstörungen	
	s1: Porropsie (Fernsehen), Nahsehen	
	s2: Mikro- und Makropsien	
	s3: Metamorphosie	
	s4: Veränderungen des Farbensehens, Farbigsehen	
	s5: Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	
	s6: Wahrnehmungsveränderungen am eigenen Gesicht	
	s7: Scheinbewegungen von Wahrnehmungsobjekten	
	s8: Doppelt-, Schief-, Schräg- und Verkehrtsehen	
	s9: Störungen der Schätzung von Entfernungen, Größe von Gegenständen	
	s10: Auflösung der Geradlinigkeit gegenständlicher Konturen	
	s11: Dysmegalopsien	
	s12: Abnorm langes Haften optischer Reize, nachträgliches Sehen	
C.2.4	Geräuschüberempfindlichkeit; Akoasmen	

BSABS-Item-Nr.	BASISSYMPTOM	?
	s1: Geräusche, Lärm, allgemein akustische Reize	
	s2: Akoasmen	
C.2.5	Veränderungen von Gehörs wahrnehmungen	
	s1: Intensität/Qualität	
	s2: Abnorm langes Haften akustischer Reize, nachträgliches Hören von Geräuschen	
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	
	s1: Olfaktorisch	
	s2: Gustatorisch	
	s3: Taktile (sensibel)	
C.2.7	Störung der Erfassung der Bedeutung von Wahrnehmungen	
C.2.8	Sensorische Überwachheit	
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	
C.2.10	Störungen der Kontinuität der Wahrnehmung der eigenen Handlungen	
C.2.11	Derealisation	
	s1: Einbuße an Physiognomierung, Entfremdung der Wahrnehmungswelt	
	s2: Zunahme der Physiognomierung	
C.3	Kognitive Handlungs- (Bewegungs-) Störungen	
C.3.1	Motorische Interferenz; Automatosesyndrom	
C.3.2	Motorische Blockierung, Bannungszustände (Starre)	
C.3.3	Verlust automatisierter Fertigkeiten (Automatismenverlust)	
C.3.4	Psychomotorische Verlangsamung, Störung der psychomotorischen Organisation der Sprache	
	s1: Verlangsamung psychomotorischer Abläufe	
	s2: Verlangsamung des Sprechens	
C.3.5	Selbst wahrgenommene Bewegungsstörungen i.S. extrapyramidal aussehender und ticartiger Hyperkinesen	
D	COENÄSTHESIEN	
D.1	Taubheits- und Steifigkeitsempfindungen	
D.1.1	Entfremdungserlebnisse am eigenen Körper - somatopsychische Depersonalisation	
D.2	Sensationen motorischer Schwäche ("Lähmungssensationen")	
D.3	Mehr umschriebene Schmerzempfindungen	
D.4	Wandersensationen	

BSABS-Item-Nr.	BASISSYMPTOM	?
D.5	Elektrisierungssensationen	
D.6	Thermische Sensationen	
D.7	Bewegungs-, Zug- und Druckempfindungen im Körperinneren und/oder an der Körperoberfläche	
D.8	Sensationen abnormer Leere, Leichtigkeit und Schwere, Fall- und Sink-, Levitations- und Elevationsphänomene	
D.9	Sensationen der Verkleinerung, Schrumpfung und Einschnürung, der Vergrößerung und Ausdehnung	
D.10	Kinästhetische Sensationen	
D.11	Sog. Vestibuläre Sensationen; qualitativ eigenartige Raumsinn- und Gleichgewichtsstörungen	
D.12	Sensorisch und sensibel ausgelöste Mißempfindungen	
D.13	Nicht rubrizierbare Coenästhesien	
D.14	Dysästhetische Krisen	
D.15	Paroxysmale (nicht psycho-reaktive, endogene) Angstzustände ohne Coen-ästhesien und ohne zentral-vegetative Störung	
E	ZENTRAL-VEGETATIVE STÖRUNGEN INCL. SCHLAFSTÖRUNGEN UND INTOLERANZ GEGEN BESTIMMTE SUBSTANZEN	
E.1	Zentral-vegetative Störungen	
E.1.1	Paroxysmen von Tachykardie oder Bradykardie	
	s1: Herzrasen	
	s2: Langsamer Herz-/Pulsschlag	
	s3: Herzaussetzen, -stillstand	
E.1.2	Vasomotorische Störungen; Störungen der Thermoregulation	
	s1: Vasomotorische Störungen	
	s2: Thermoregulation (v.a. Kältezittern)	
E.1.3	Übelkeit, Brechreiz und Erbrechen; Aufstoßen	
E.1.4	Veränderungen einzelner Vitaltriebe	
	s1: Appetitlosigkeit	
	s2: Heißhunger	
	s3: Appetenzwandel	
	s4: Durstempfindung	
	s5: Suchtähnlicher Nikotin- und Alkoholabusus	
	s6: Ostipation, Diarrhoe	
E.1.5	Veränderung von Libido und Potenz, Menstruationsstörungen	
	s1: Minderung der Libido	

BSABS-Item-Nr.	BASISSYMPTOM	?
	s2: Steigerung der Libido	
	s3: Verlust der Potenz	
	s4: Menstruationsstörungen	
E.1.6	Störungen der Speichel-, Schweiß- und/oder Talgdrüsensekretion	
	s1: Speichel	
	s2: Schweiß	
	s3: Talg	
	s4: 'Symptom der widerspenstigen Haare'	
E.1.7	Polyurie, Nykturie, Oligurie, Urininkontingenz/-retention; Harn- und Stuhlzwang	
E.1.8	Tachypnoe (Polypnoe, Dyspnoe)	
E.2	Schlafstörungen	
E.2.1	Einschlafstörungen	
E.2.2	Durchschlafstörungen	
E.2.3	Durchschlafstörungen i.S. von Früherwachen	
E.2.4	Kombinierte Einschlaf- und Durchschlafstörungen	
E.2.5	Schlafinversion (-umkehr). Abnorm tiefer und langer Schlaf	
E.3	Intoleranz gegen Alkohol, Coffein, Nikotin u.a. Substanzen	
E.3.1	Intoleranz gegen Alkohol	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
E.3.2	Intoleranz gegen Coffein (Kaffee, Tee)	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
E.3.3	Intoleranz gegen Nikotin	
	IMS-1: Innere Erregung, Unruhe	

BSABS-Item-Nr.	BASISSYMPTOM	?
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
E.3.4	Intoleranz gegen bestimmte Speisen und Getränke oder bestimmte andere Substanzen	
	IMS-1: Innere Erregung, Unruhe	
	IMS-2: Schlafstörungen	
	IMS-3: Zwanghaftes Grübeln	
	IMS-4: Coenästhesien	
	IMS-5: Zentral-vegetative Störungen	
	IMS-6: Konzentrationsstörungen	
F	BEWÄLTIGUNGSVERSUCHE	
F.1	Vermeidungsverhalten	
F.2	Bestreben, die selbst wahrgenommenen Basissymptome durch bestimmte Verhaltensweisen zu kompensieren	
F.3	Gewöhnung und/oder Anpassung an die Erkrankung	
F.4	Bemühungen, die Basissymptome durch willensmäßige Anstrengung zu kompensieren	
F.5	Versuch, bestimmte Funktionen oder Verhaltensweisen zu trainieren	
F.6	'Selbstbehandlung' mit Alkohol, Medikamenten u.ä.	

- ?: 2 = Basissymptom vorhanden;
1 = Basissymptom fraglich vorhanden;
0 = Basissymptom nicht vorhanden

ANHANG B: Aachener Kurzversion der BSABS: Erhebungsbogen der 66 Items

BSABS-Item-Nr.	BASISSYMPTOM	?
A / B	DYNAMISCHE DEFIZIENZEN	
A.6.1	Veränderungen von Grundstimmung und emotionaler Resonanzfähigkeit	
A.6.2	Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten	
A.6.4	Minderung des Kontaktbedürfnisses	
A.7.1	Minderung der Kontaktfähigkeit bei vorhandenem Kontaktwunsch	
A.7.2	Störung des In-Erscheinung-Tretens	
A.8.1 + B.1.2	Minderung der psychischen Belastungsfähigkeit gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen	
A.8.2 + B.1.3	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen	
A.8.3 + B.1.4	Minderung der psychischen Belastungsfähigkeit gegenüber Arbeit unter Zeitdruck und/oder rasch wechselnden unterschiedlichen Anforderungen	
B.1.1	Minderung der psychischen Belastungsfähigkeit gegenüber körperlicher und/oder psychischer arbeitsmäßiger Beanspruchung	
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	
B.2.1	Erhöhte Beeindruckbarkeit durch alltägliche Ereignisse	
B.2.2	Erhöhte Beeindruckbarkeit durch Verhaltensweisen anderer, die die Person persönlich betreffen	
B.2.3	Erhöhte Beeindruckbarkeit durch 'fremdes Leid'	
C.1	Kognitive Denkstörungen	
C.1.1	Gedankeninterferenz	
C.1.2	s1: Zwangähnliches Perseverieren zurückliegender Vorgänge	
C.1.3	Gedankendrängen, Gedankenjagen	
C.1.4	Blockierung des jeweiligen Gedankenganges	
C.1.5	Störung der Konzentrationsfähigkeit	
C.1.6	Störung der rezeptiven Sprache	
C.1.7	Störung der expressiven Sprache	
C.1.8	Störung des unmittelbaren Behaltens	
C.1.9	Störung des Kurzzeitgedächtnisses	
C.1.10	Besonders strukturierte Störungen des Langzeitgedächtnisses	
C.1.12	Verlangsamung und Erschwerung der Denkvorgänge	
C.1.13	Störung der Denkinitiative und gedankliche Intentionalität	
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	

BSABS-Item-Nr.	BASISSYMPTOM	?
C.1.16	Störung der Symbolerfassung (Konkretismus)	
C.1.17	'Subjekt-Zentrismus' - Eigenbeziehungstendenz	
C.2	Kognitive Wahrnehmungsstörungen	
C.2.1	s1: Verschwommen-, Trübsehen	
C.2.1	s3: Partielles Sehen	
C.2.2	s1: Überempfindlichkeit gegen Licht oder visuelle Reize	
C.2.2	s2: Photopsien	
C.2.3	s1: Porropsie (Fernsehen), Nahsehen	
C.2.3	s2: Mikro- und Makropsien	
C.2.3	s3: Metamorphosie	
C.2.3	s4: Veränderungen des Farbensehens, Farbigsehen	
C.2.3	s5: Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	
C.2.3	s6: Wahrnehmungsveränderungen am eigenen Gesicht	
C.2.3	s7: Scheinbewegungen von Wahrnehmungsobjekten	
C.2.3	s8: Doppelt-, Schief-, Schräg- und Verkehrtsehen	
C.2.3	s9: Störungen der Schätzung von Entfernungen, Größe von Gegenständen	
C.2.3	s10: Auflösung der Geradlinigkeit gegenständlicher Konturen	
C.2.3	s12: Abnorm langes Haften optischer Reize, nachträgliches Sehen	
C.2.4	s1: Geräusche, Lärm, allgemein akustische Reize	
C.2.4	s2: Akoasmen	
C.2.5	s1: Intensität/Qualität	
C.2.5	s2: Abnorm langes Haften akustischer Reize, nachträgliches Hören	
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	
C.2.8	Sensorische Überwachheit	
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	
C.3	Kognitive Handlungs- (Bewegungs-) Störungen	
C.3.1	Motorische Interferenz; Automatosesyndrom	
C.3.2	Motorische Blockierung, Bannungszustände (Starre)	
C.3.3	Verlust automatisierter Fertigkeiten (Automatismenverlust)	
D	COENÄSTHESIEN	
D.1	Taubheits- und Steifigkeitsempfindungen	

BSABS-Item-Nr.	BASISSYMP TOM	?
D.1.1	Entfremdungserlebnisse am eigenen Körper - somatopsychische Depersonalisation	
D.2	Sensationen motorischer Schwäche ("Lähmungssensationen")	
D.3	Mehr umschriebene Schmerzempfindungen	
D.4	Wandersensationen	
D.5	Elektrisierungssensationen	
D.6	Thermische Empfindungen	
D.7	Bewegungs-, Zug- und Druckempfindungen im Körperinneren und/oder an der Körperoberfläche	
D.8	Sensationen abnormer Leere, Leichtigkeit und Schwere, Fall- und Sink-, Levitations- und Elevationsphänomene	
D.9	Sensationen der Verkleinerung, Schrumpfung und Einschnürung, der Vergrößerung und Ausdehnung	
D.10	Kinästhetische Empfindungen	
D.11	Sog. Vestibuläre Empfindungen; qualitativ eigenartige Raumsinn- und Gleichgewichtsstörungen	
D.14	Dysästhetische Krisen	

- ?: 2 = Basissymptom vorhanden;
1 = Basissymptom fraglich vorhanden;
0 = Basissymptom nicht vorhanden

ANHANG C: 72-Item Kurzversion der BSABS: Erhebungsbogen

BSABS-Item-Nr.	BASISSYMPTOM	?
A / B	DYNAMISCHE DEFIZIENZEN	
A.5	Mangelnde Entscheidungsfähigkeit, Entschlußschwäche, Unschlüssigkeit	
A.6.1	Veränderungen von Grundstimmung und emotionaler Resonanzfähigkeit	
A.6.2	Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten	
A.6.3	Abschwächung bejahender Fremdwert- und Sympathiegefühle	
A.6.4	Minderung des Kontaktbedürfnisses	
A.7.1	Minderung der Kontaktfähigkeit bei vorhandenem Kontaktwunsch	
A.7.2	Störung des In-Erscheinung-Tretens	
A.8.1 + B.1.2	Minderung der psychischen Belastungsfähigkeit gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen	
A.8.2 + B.1.3	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen	
A.8.3 + B.1.4	Minderung der psychischen Belastungsfähigkeit gegenüber Arbeit unter Zeitdruck und/oder rasch wechselnden unterschiedlichen Anforderungen	
B.1.1	Minderung der psychischen Belastungsfähigkeit gegenüber körperlicher und/oder psychischer arbeitsmäßiger Beanspruchung	
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	
B.2.1	Erhöhte Beeindruckbarkeit durch alltägliche Ereignisse	
B.2.2	Erhöhte Beeindruckbarkeit durch Verhaltensweisen anderer, die die Person persönlich betreffen	
B.2.3	Erhöhte Beeindruckbarkeit durch 'fremdes Leid'	
B.3.1	Erhöhte Reflexivität: Verlust an Naivität, Unbekümmertheit, Unbefangtheit	
B.3.4	Autopsychische Depersonalisation	
C.1	Kognitive Denkstörungen	
C.1.1	Gedankeninterferenz	
C.1.2	s1: Zwangähnliches Perseverieren zurückliegender Vorgänge	
C.1.3	Gedankendrängen, Gedankenjagen	
C.1.4	Blockierung des jeweiligen Gedankenganges	
C.1.5	Störung der Konzentrationsfähigkeit	
C.1.6	Störung der rezeptiven Sprache	
C.1.7	Störung der expressiven Sprache	
C.1.8	Störung des unmittelbaren Behaltens	
C.1.9	Störung des Kurzzeitgedächtnisses	

BSABS-Item-Nr.	BASISSYMPTOM	?
C.1.10	Besonders strukturierte Störungen des Langzeitgedächtnisses	
C.1.12	Verlangsamung und Erschwerung der Denkvorgänge	
C.1.13	Störung der Denkinitiative und gedankliche Intentionalität	
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	
C.1.16	Störung der Symbolerfassung (Konkretismus)	
C.1.17	'Subjekt-Zentrismus' - Eigenbeziehungstendenz	
C.2	Kognitive Wahrnehmungsstörungen	
C.2.1	s1: Verschwommen-, Trübsehen	
C.2.1	s3: Partielles Sehen	
C.2.2	s1: Überempfindlichkeit gegen Licht oder visuelle Reize	
C.2.2	s2: Photopsien	
C.2.3	s1: Porropsie (Fernsehen), Nahsehen	
C.2.3	s2: Mikro- und Makropsien	
C.2.3	s3: Metamorphosie	
C.2.3	s4: Veränderungen des Farbsehens, Farbigsehen	
C.2.3	s5: Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	
C.2.3	s6: Wahrnehmungsveränderungen am eigenen Gesicht	
C.2.3	s7: Scheinbewegungen von Wahrnehmungsobjekten	
C.2.3	s8: Doppelt-, Schief-, Schräg- und Verkehrtsehen	
C.2.3	s9: Störungen der Schätzung von Entfernungen, Größe von Gegenständen	
C.2.3	s10: Auflösung der Geradlinigkeit gegenständlicher Konturen	
C.2.3	s12: Abnorm langes Haften optischer Reize, nachträgliches Sehen	
C.2.4	s1: Geräusche, Lärm, allgemein akustische Reize	
C.2.4	s2: Akoasmen	
C.2.5	s1: Intensität/Qualität	
C.2.5	s2: Abnorm langes Haften akustischer Reize, nachträgliches Hören von Geräuschen	
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	
C.2.8	Sensorische Überwachheit	
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	
C.2.10	Störungen der Kontinuität der Wahrnehmung der eigenen Handlungen	
C.2.11	Derealisation	

BSABS-Item-Nr.	BASISSYMP TOM	?
C.3	Kognitive Handlungs- (Bewegungs-) Störungen	
C.3.1	Motorische Interferenz; Automatosesyndrom	
C.3.2	Motorische Blockierung, Bannungszustände (Starre)	
C.3.3	Verlust automatisierter Fertigkeiten (Automatismenverlust)	
D	COENÄSTHESIEN	
D.1	Taubheits- und Steifigkeitsempfindungen	
D.1.1	Entfremdungserlebnisse am eigenen Körper - somatopsychische Depersonalisation	
D.2	Sensationen motorischer Schwäche ("Lähmungssensationen")	
D.3	Mehr umschriebene Schmerzempfindungen	
D.4	Wandersensationen	
D.5	Elektrisierungssensationen	
D.6	Thermische Sensationen	
D.7	Bewegungs-, Zug- und Druckempfindungen im Körperinneren und/oder an der Körperoberfläche	
D.8	Sensationen abnormer Leere, Leichtigkeit und Schwere, Fall- und Sink-, Levitations- und Elevationsphänomene	
D.9	Sensationen der Verkleinerung, Schrumpfung und Einschnürung, der Vergrößerung und Ausdehnung	
D.10	Kinästhetische Sensationen	
D.11	Sog. Vestibuläre Sensationen; qualitativ eigenartige Raumsinn- und Gleichgewichtsstörungen	
D.14	Dysästhetische Krisen	

?: 2 = Basissymptom vorhanden;
1 = Basissymptom fraglich vorhanden;
0 = Basissymptom nicht vorhanden

ANHANG D: Beurteilungsbogen für Entscheidungsregeln

Im folgenden sind acht Varianten zur Vorhersage von schizophrenen Ersterkrankungen aufgeführt. Bitte schauen Sie sich genau an und beurteilen Sie sie auf der Grundlage Ihrer bisherigen klinischen Erfahrungen im Hinblick auf zwei Fragen:

1. Wie plausibel erscheinen die Lösungen inhaltlich für die Vorhersage einer schizophrenen Erstepisode?
2. Wenn Sie selbst eine der Lösungen in der klinischen Praxis verwenden müssten, welche würden Sie auswählen?

Bitte bringen Sie die Lösungen in eine entsprechende Rangordnung mit

Plausibilität: 1=plausibelste Lösung, 2=zweitplausibelste Lösung,, 8=am wenigsten plausible Lösung

Anwendbarkeit: 1=meine erste Wahl, 2=meine zweite Wahl,, 8=meine letzte Wahl

Beachte: Basieren die Vorhersagealgorithmen auf mathematischen Formeln, so nimmt ein **Basissymptom** den Wert **1** an, wenn es **vorhanden** ist, und den Wert **0**, wenn es **nicht vorhanden** ist.

Vorhersagealgorithmus	Plausibilität	Anwendbarkeit
Die Wahrscheinlichkeit für eine späteren schizophrenen Erkrankung beträgt $1/1+e^{(-1 \times (0.9682 \times [\text{die Anzahl der von den folgenden 9 Symptomen vorhandenen: 'Unfähigkeit, die Aufmerksamkeit zu spalten', 'Gedankeninterferenz', 'Gedankendrängen, -jagen', 'Gedankenblockierung', 'Störung der rezeptiven Sprache', 'Störung der expressiven Sprache', 'Störung der Symbolerfassung', 'Eigenbeziehungstendenz', 'Fesselung durch Wahrnehmungsdetails'] - 1,4569))}$. Ist die Wahrscheinlichkeit größer als 0,5 wird eine spätere schizophrene Erkrankung vorhergesagt.		
Eine schizophrene Erkrankung wird vorhergesagt, wenn mindestens 2 der folgenden 8 Symptome vorhanden sind: 'Unfähigkeit, die Aufmerksamkeit zu spalten', 'Gedankeninterferenz', 'zwangähnliches Perseverieren von Denkinhalten', 'Gedankendrängen, -jagen', 'Gedankenblockierung', 'Störung der rezeptiven Sprache', 'Eigenbeziehungstendenz', 'De-realisation'.		
Wenn entweder 'Gedankeninterferenz', 'zwangähnliches Perseverieren von Denkinhalten', 'Gedankendrängen, -jagen' und/oder 'Störung der rezeptiven Sprache' vorhanden ist, wird eine schizophrene Erkrankung vorhergesagt. Ist keines dieser 4 Symptome vorhanden, wird eine schizophrene Entwicklung ausgeschlossen.		

Vorhersagealgorithmus	Plausibilität	Anwendbarkeit
<p>Eine schizophrene Erkrankung wird vorhergesagt, wenn mindestens 3 der folgenden 17 Symptome vorhanden sind: 'Unfähigkeit, die Aufmerksamkeit zu spalten', 'Gedankeninterferenz', 'Gedankendrängen, -jagen', 'Störung der rezeptiven Sprache', 'Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. Phantasie- und Erinnerungsvorstellungen', 'Eigenbeziehungstendenz', 'Lichtüberempfindlichkeit', 'Photopsien', 'Porropsie, Nahsehen', 'Mikropsie, Makropsie', 'Wahrnehmungsveränderungen am eigenen Gesicht', 'Scheinbewegungen von Wahrnehmungsobjekten', 'Störung der Schätzung von Entfernungen und Größen von Gegenständen', 'abnorm langes Haften optischer Reize', 'Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder taktilem Gebiet', 'sensorische Überwachheit', 'Verlust automatisierter Fertigkeiten'.</p>		
<p>Eine spätere schizophrene Erkrankung wird vorhergesagt, wenn der Diskriminanzwert [= konstantes Glied + Diskriminanzkoeffizient x (die Anzahl der von den folgenden 9 Symptomen vorhandenen: 'Unfähigkeit, die Aufmerksamkeit zu spalten', 'Gedankeninterferenz', 'Gedankendrängen, -jagen', 'Gedankenblockierung', 'Störung der rezeptiven Sprache', 'Störung der expressiven Sprache', 'Störung der Symbolerfassung', 'Eigenbeziehungstendenz', 'Fesselung durch Wahrnehmungsdetails')] näher an dem Gruppencentroid von 0,723 als am Gruppencentroid von -0,612 liegt.</p>		
<p>Eine spätere schizophrene Erkrankung wird vorhergesagt, wenn der Diskriminanzwert [= konstantes Glied + Diskriminanzkoeffizient1 x 'zwangähnliches Perseverieren von Denkinhalten' + Diskriminanzkoeffizient2 x 'Gedankendrängen, -jagen' + Diskriminanzkoeffizient3 x 'Störung der rezeptiven Sprache' + Diskriminanzkoeffizient4 x 'Wahrnehmungsveränderungen am eigenen Gesicht'] im 4-dimensionalen Raum näher an dem Gruppencentroid der später schizophrenen Gruppe als am Gruppencentroid der später nicht schizophrenen Gruppe liegt.</p>		
<p>Die Wahrscheinlichkeit für eine späteren schizophrenen Erkrankung beträgt $1/1+e^{(-1 \times (1,2951 \times \text{'zwangähnliches Perseverieren von Denkinhalten'} + 2,0743 \times \text{'Gedankendrängen, -jagen'} + 2,2148 \times \text{'Störung der rezeptiven Sprache'} + 1,7343 \times \text{'Wahrnehmungsveränderungen am eigenen Gesicht'} - 1,4402))}$. Ist die Wahrscheinlichkeit größer als 0,5 wird eine spätere schizophrene Erkrankung vorhergesagt.</p>		

Vorhersagealgorithmus	Plausibilität	Anwendbarkeit
<p>Eine schizophrene Erkrankung wird vorhergesagt, wenn ‘Gedankeninterferenz’ vorhanden und ‘Gedankendrängen, -jagen’ vorhanden oder ‘Gedankendrängen, -jagen’ vorhanden und ‘Gedankenblockierungen’ vorhanden oder ‘Gedankeninterferenz’ vorhanden und ‘Gedankenblockierungen’ nicht vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘zwangähnliches Perseverieren von Denkinhalten’ nicht vorhanden und ‘Störung der rezeptiven Sprache’ vorhanden oder ‘zwangähnliches Perseverieren von Denkinhalten’ nicht vorhanden und ‘Eigenbeziehungstendenz’ vorhanden oder ‘zwangähnliches Perseverieren von Denkinhalten’ vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘Gedankeninterferenz’ nicht vorhanden und ‘zwangähnliches Perseverieren von Denkinhalten’ vorhanden</p> <p>Eine schizophrene Entwicklung wird ausgeschlossen, wenn ‘Gedankeninterferenz’ vorhanden und ‘zwangähnliches Perseverieren von Denkinhalten’ vorhanden und ‘Gedankenblockierungen’ nicht vorhanden und ‘Störung der rezeptiven Sprache’ nicht vorhanden und ‘Eigenbeziehungstendenz’ vorhanden oder ‘Gedankendrängen, -jagen’ nicht vorhanden und ‘Gedankenblockierungen’ vorhanden und ‘Störung der rezeptiven Sprache’ vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘Gedankeninterferenz’ vorhanden und ‘zwangähnliches Perseverieren von Denkinhalten’ vorhanden und ‘Gedankendrängen, -jagen’ nicht vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘Gedankeninterferenz’ nicht vorhanden und ‘Gedankendrängen, -jagen’ nicht vorhanden und ‘Gedankenblockierungen’ nicht vorhanden und ‘Störung der rezeptiven Sprache’ nicht vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘zwangähnliches Perseverieren von Denkinhalten’ nicht vorhanden und ‘Gedankendrängen, -jagen’ nicht vorhanden und ‘Störung der rezeptiven Sprache’ nicht vorhanden und ‘Eigenbeziehungstendenz’ nicht vorhanden oder ‘Gedankeninterferenz’ nicht vorhanden und ‘zwangähnliches Perseverieren von Denkinhalten’ nicht vorhanden und ‘Gedankenblockierungen’ nicht vorhanden und ‘Störung der rezeptiven Sprache’ nicht vorhanden</p>		

ANHANG E: Ergebnis der Häufigkeitsvergleiche der Einzelvariablen in beiden Outcome-Gruppen der Berechnungsstichprobe (n = 120).

BSABS-Item-Nr.	Basissymptom	Häufigkeit in nicht übergegangener Gruppe	Häufigkeit in übergegangener Gruppe	χ^2 -Wert ^a	p ^b
A.5	Mangelnde Entscheidungsfähigkeit	17	10	1,086	0,297
A.6.1	Veränderungen von Grundstimmung und emotionaler Resonanzfähigkeit	42	35	0,012	0,911
A.6.2	Unfähigkeit zur Diskriminierung verschiedener Gefühlsqualitäten	4	7	1,546	0,214
A.6.3	Abschwächung bejahender Fremdwert- und Sympathiegefühle	32	27	0,000	0,988
A.6.4	Minderung des Kontaktbedürfnisses	20	19	0,194	0,660
A.7.1	Minderung der Kontaktfähigkeit bei vorhandenem Kontaktwunsch	19	13	0,477	0,490
A.7.2	Störung des In-Erscheinung-Tretens	9	12	1,311	0,252
A.8.1 + B.1.2	Minderung der psychischen Belastungsfähigkeit gegenüber ungewöhnlichen, unerwarteten, besonderen, neuen Anforderungen	15	21	3,237	0,072
A.8.2 + B.1.3	Minderung der psychischen Belastungsfähigkeit gegenüber bestimmten alltäglichen, primär affektiv neutralen sozialen Situationen	27	23	0,001	0,975
A.8.3 + B.1.4	Minderung der psychischen Belastungsfähigkeit gegenüber Arbeit unter Zeitdruck oder rasch wechselnden unterschiedlichen Anforderungen	21	15	0,360	0,549
B.1.1	Minderung der psychischen Belastungsfähigkeit gegenüber körperlicher und/oder psychischer arbeitsmäßiger Beanspruchung	26	22	0,000	1,000
A.8.4	Unfähigkeit, die Aufmerksamkeit zu spalten	4	14	8,704	0,003
B.2.1	Erhöhte Beeindruckbarkeit durch alltägliche Ereignisse	8	6	0,057	0,812
B.2.2	Erhöhte Beeindruckbarkeit durch Verhaltensweisen anderer, die die Person persönlich betreffen	18	15	0,003	0,959
B.2.3	Erhöhte Beeindruckbarkeit durch 'fremdes Leid'	7	6	0,001	0,980
B.3.1	Erhöhte Reflexivität	16	16	0,305	0,581
B.3.4	Autopsychische Depersonalisation	9	10	0,420	0,517

BSABS-Item-Nr.	Basissymptom	Häufigkeit in nicht übergegangener Gruppe	Häufigkeit in übergegangener Gruppe	χ^2-Wert ^a	p ^b
C.1.1	Gedankeninterferenz	6	24	18,808	0,000
C.1.2	Zwangähnliches Perseverieren zurückliegender Vorgänge	8	19	8,449	0,004
C.1.3	Gedankendrängen, -jagen	3	21	20,979	0,000
C.1.4	Blockierung des jeweiligen Gedankengangs	9	20	8,242	0,004
C.1.5	Störung der Konzentrationsfähigkeit	43	43	2,123	0,145
C.1.6	Störung der rezeptiven Sprache	4	22	20,108	0,000
C.1.7	Störung der expressiven Sprache	5	13	5,940	0,015
C.1.8	Störung des unmittelbaren Behaltens	10	8	0,016	0,898
C.1.9	Störung des Kurzzeitgedächtnisses	12	12	0,210	0,647
C.1.10	Besonders strukturierte Störungen des Langzeitgedächtnisses	4	3	0,027	1,000
C.1.12	Verlangsamung und Erschwerung der Denkvorgänge	19	23	2,075	0,150
C.1.13	Störung der Denkinitiative und gedanklichen Intentionalität	16	14	0,011	0,916
C.1.15	Störung der Diskriminierung von Vorstellungen und Wahrnehmungen bzw. von Phantasie- und Erinnerungsvorstellungen	3	13	9,328	0,002
C.1.16	Störung der Symbolerfassung	0	3	3,636	0,057
C.1.17	'Subjekt-Zentrismus', Eigenbeziehungstendenz	6	21	14,320	0,000
C.2.1 s1	Verschwommen-, Trübsehen	5	8	1,448	0,229
C.2.1 s3	Partielles Sehen	0	3	3,636	0,093
C.2.2 s1	Überempfindlichkeit gegen Licht oder visuelle Reize	2	12	10,154	0,001
C.2.2 s2	Photopsien	2	9	6,316	0,012
C.2.3 s1	Fern-, Nahsehen	1	2	0,538	0,593
C.2.3 s2	Mikro- und Makropsien	1	3	1,418	0,332
C.2.3 s3	Metamorphopsie	0	1	1,192	0,458
C.2.3 s4	Veränderungen des Farbensehens	2	5	1,962	0,245
C.2.3 s5	Wahrnehmungsveränderungen an Gesicht und/oder Gestalt anderer	0	7	8,785	0,003
C.2.3 s6	Wahrnehmungsveränderungen am eigenen Gesicht	3	12	8,061	0,005
C.2.3 s7	Scheinbewegungen von Wahrnehmungsobjekten	1	3	1,418	0,332

BSABS-Item-Nr.	Basissymptom	Häufigkeit in nicht übergegangener Gruppe	Häufigkeit in übergegangener Gruppe	χ^2 -Wert ^a	p ^b
C.2.3 s8	Doppelt-, Schief-, Schräg- und Verkehrtsehen	0	1	1,192	0,458
C.2.3 s9	Störungen der Schätzung von Entfernungen, Größe von Gegenständen	1	2	0,538	0,593
C.2.3 s10	Auflösung der Geradlinigkeit gegenständlicher Figuren	0	0	c	c
C.2.3 s12	Abnorm langes Haften optischer Reize, nachträgliches Sehen	1	2	0,538	0,593
C.2.4 s1	Geräuschüberempfindlichkeit	13	15	0,881	0,348
C.2.4 s2	Akoasmen	5	8	1,448	0,229
C.2.5 s1	Veränderung der Intensität/Qualität von Gehörs wahrnehmungen	5	11	3,905	0,048
C.2.5 s2	Abnorm langes Haften akustischer Reize, nachträgliches Hören	0	1	1,192	0,458
C.2.6	Wahrnehmungsveränderungen auf olfaktorischem, gustatorischem oder sensiblem Gebiet	4	13	7,488	0,006
C.2.8	Sensorische Überwachtheit	1	11	11,282	0,001
C.2.9	Fesselung (Bannung) durch Wahrnehmungsdetails	0	7	8,785	0,003
C.2.10	Störung der Kontinuität der Wahrnehmung der eigenen Handlungen	2	3	0,422	0,660
C.2.11	Derealisation	7	14	4,450	0,035
C.3.1	Motorische Interferenz	0	6	7,464	0,008
C.3.2	Motorische Blockierung, Bannungs Zustände	4	9	3,215	0,073
C.3.3	Automatismenverlust	3	12	8,061	0,005
D.1	Taubheits-, Steifigkeitsempfindungen	12	8	0,329	0,566
D.1 s1	Somatopsychische Depersonalisation	9	10	0,420	0,517
D.2	Sensationen motorischer Schwäche	9	6	0,235	0,628
D.3	Mehr umschriebene Schmerzempfindungen	21	10	3,103	0,078
D.4	Wandersensationen	11	4	2,537	0,111
D.5	Elektrisierungssensationen	6	3	0,612	0,505
D.6	Thermische Sensationen	16	10	0,727	0,394
D.7	Bewegungs-, Zug- und Druckempfindungen im Körperinneren und/oder an der Körperoberfläche	21	26	2,800	0,094

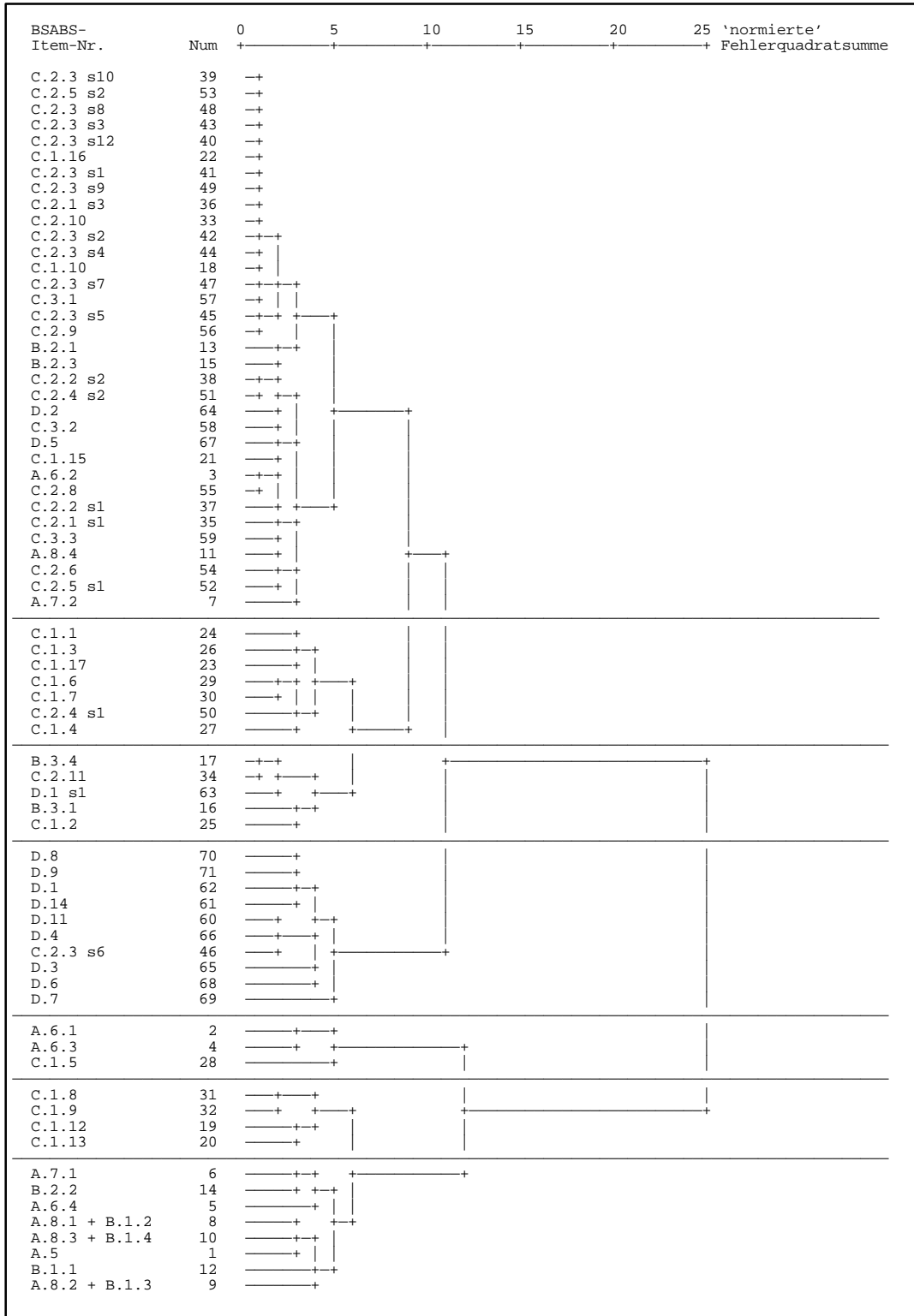
BSABS-Item-Nr.	Basissymptom	Häufigkeit in nicht übergegangener Gruppe	Häufigkeit in übergegangener Gruppe	χ^2 -Wert ^a	p ^b
D.8	Sensationen abnormer Leere, Leichtigkeit und Schwere, Fall- und Sink-, Levitations- und Elevationsphänomene	8	13	2,648	0,104
D.9	Sensationen der Verkleinerung, Schrumpfung und Einschnürung, der Vergrößerung und Ausdehnung	14	12	0,001	0,970
D.10	Kinästhetische Sensationen	0	0	^c	^c
D.11	Vestibuläre Sensationen	12	9	0,091	0,763
D.14	Dysästhetische Krisen	20	15	0,176	0,675

^a 2x2- χ^2 -Test, df = 1

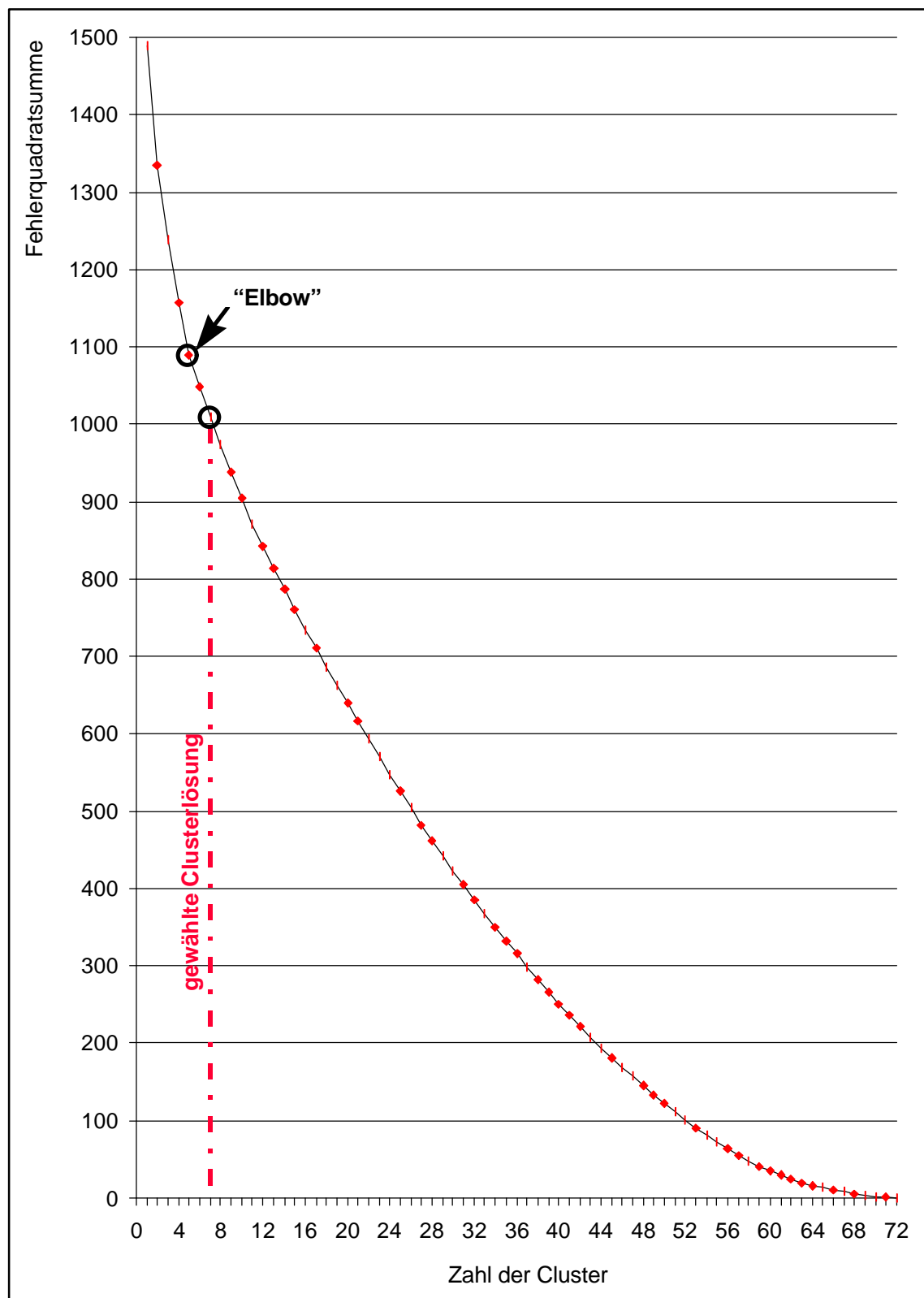
^b Asymptotische bzw. exakte Signifikanz bei Zellenbesetzungen unterhalb der erwarteten Häufigkeit von $f_{(e)} \geq 5$ bei zweiseitiger Testung

^c Kann nicht berechnet werden, da nur zwei Zellen besetzt sind.

ANHANG F: Hierarchische Clusteranalyse nach der Wald-Linkage-Methode: Dendrogramm der Fusionierungsstufen und der 'normierten' Fehlerquadratsummen (d.h. der höchste Wert der Fehlerquadratsumme bei der letzten Fusionierungsstufe ist auf 25 festgelegt)



Anhang G: Entwicklung der Fehlerquadratsumme: 'Elbow'-Kriterium zur Bestimmung der Clusteranzahl



ANHANG H: Korrelationsmatrix zwischen Einzelvariablen und Clustern nach der letzten, zweiten Umstellung: Pearsons Korrelationskoeffizient.

BSABS-Item-Nr.	Cluster						
	BS1	BS2	BS3	BS4	BS5	BS6	BS7
A.5	0,551 **	0,196 *	0,062	0,146	0,122	0,033	0,120
A.6.1	0,222 *	0,792 **	0,123	0,318 **	0,264 **	0,124	-0,037
A.6.2	0,201 *	0,216 *	0,449 **	0,513 **	0,247 **	0,239 **	0,143
A.6.3	0,138	0,764 **	0,129	0,316 **	0,190 *	0,068	-0,015
A.6.4	0,444 **	0,328 **	0,154	0,268 **	0,316 **	0,249 **	-0,089
A.7.1	0,487 **	0,273 **	0,109	0,131	0,247 **	0,168	-0,244 **
A.7.2	0,347 **	-0,040	0,196 *	0,214 *	0,117	0,109	0,070
A.8.1 + B.1.2	0,754 **	0,149	0,186 *	0,266 **	0,310 **	0,211 *	-0,010
A.8.2 + B.1.3	0,624 **	0,074	0,225 *	0,074	0,246 **	0,312 **	0,068
A.8.3 + B.1.4	0,670 **	0,167	0,145	0,110	0,338 **	0,102	-0,054
A.8.4	0,421 **	0,199 *	0,446 **	0,294 **	0,290 **	0,552 **	0,150
B.1.1	0,509 **	0,070	0,065	0,032	0,208 *	0,057	0,101
B.2.1	0,484 **	0,155	0,119	0,249 **	0,059	0,095	-0,061
B.2.2	0,656 **	0,182 *	0,082	0,042	0,142	0,060	-0,161
B.2.3	0,468 **	0,077	0,101	0,170	0,040	0,004	-0,016
B.3.1	0,371 **	0,273 **	0,257 **	0,556 **	0,218 *	0,178	0,062
B.3.4	0,112	0,265 **	0,345 **	0,770 **	0,212 *	0,274 **	0,125
C.1.1	0,110	0,160	0,324 **	0,277 **	0,132	0,620 **	-0,002
C.1.2	0,150	0,235 **	0,271 **	0,583 **	0,075	0,152	0,071
C.1.3	0,042	0,094	0,394 **	0,226 *	0,029	0,627 **	-0,002
C.1.4	0,073	0,197 *	0,413 **	0,153	0,225 *	0,558 **	0,151
C.1.5	0,235 **	0,577 **	0,159	0,204 *	0,207 *	0,296 **	-0,003
C.1.6	0,222 *	0,097	0,462 **	0,190 *	0,266 **	0,739 **	0,017
C.1.7	0,213 *	0,085	0,307 **	0,155	0,254 **	0,615 **	-0,058
C.1.8	0,312 **	0,039	0,194 *	0,124	0,675 **	0,160	-0,077
C.1.9	0,211 *	0,196 *	0,129	0,268 **	0,765 **	0,142	-0,036
C.1.10	0,091	0,106	0,128	0,171	0,305 **	0,084	-0,078
C.1.12	0,296 **	0,347 **	0,214 *	0,190 *	0,776 **	0,258 **	0,033
C.1.13	0,273 **	0,254 **	0,266 **	0,252 **	0,661 **	0,224 *	-0,041
C.1.15	0,052	0,106	0,465 **	0,264 **	0,094	0,275 **	0,194 *
C.1.16	-0,056	0,076	0,174	0,059	0,250 **	0,331 **	-0,044
C.1.17	0,119	0,118	0,301 **	0,199 *	-0,004	0,508 **	-0,010
C.2.1 s1	0,002	0,077	0,481 **	0,295 **	0,250 **	0,309 **	0,289 **

BSABS-Item-Nr.	Cluster						
	BS1	BS2	BS3	BS4	BS5	BS6	BS7
C.2.1 s3	0,294 **	0,180 *	0,612 **	0,413 **	0,250 **	0,331 **	0,129
C.2.2 s1	0,234 *	0,079	0,555 **	0,266 **	0,099	0,306 **	0,276 **
C.2.2 s2	0,134	0,103	0,513 **	0,302 **	-0,002	0,349 **	0,132
C.2.3 s1	0,212 *	0,128	0,453 **	0,413 **	0,166	0,129	0,194 *
C.2.3 s2	0,240 **	0,164	0,554 **	0,356 **	0,217 *	0,249 **	0,163
C.2.3 s3	0,074	0,103	0,316 **	-0,007	0,071	0,272 **	-0,013
C.2.3 s4	0,064	0,106	0,447 **	0,194 *	0,138	0,334 **	0,153
C.2.3 s5	0,201 *	0,071	0,540 **	0,123	0,054	0,488 **	0,196 *
C.2.3 s6	-0,055	-0,092	0,193 *	0,056	-0,062	0,235 **	0,397 **
C.2.3 s7	-0,047	0,073	0,242 **	0,263 **	-0,001	0,099	0,068
C.2.3 s8	-0,103	0,013	-0,026	-0,007	-0,073	-0,026	0,136
C.2.3 s9	0,212 *	0,128	0,453 **	0,448 **	0,166	0,129	0,238 **
C.2.3 s10	a	a	a	a	a	a	a
C.2.3 s12	0,150	0,180 *	0,233 *	0,519 **	0,418 **	0,157	0,086
C.2.4 s1	0,118	0,139	0,556 **	0,287 **	0,166	0,432 **	0,275 **
C.2.4 s2	-0,039	-0,028	0,461 **	0,135	0,061	0,164	0,311 **
C.2.5 s1	0,213 *	0,034	0,565 **	0,232 *	0,228 *	0,381 **	0,244 **
C.2.5 s2	0,180 *	0,103	0,214 *	0,418 **	0,287 **	0,173	-0,013
C.2.6	0,107	0,060	0,482 **	0,272 **	0,147	0,236 **	0,235 **
C.2.8	0,259 **	0,239 **	0,724 **	0,381 **	0,346 **	0,508 **	0,236 **
C.2.9	-0,019	0,071	0,234 *	0,147	0,026	0,450 **	0,052
C.2.10	0,281 **	0,153	0,439 **	0,206 *	0,359 **	0,258 **	0,106
C.2.11	0,067	0,261 **	0,327 **	0,693 **	0,169	0,275 **	0,194 *
C.3.1	0,038	0,071	0,363 **	0,085	0,118	0,350 **	0,030
C.3.2	-0,060	-0,080	0,371 **	0,099	-0,107	0,178	0,311 **
C.3.3	0,101	0,179	0,504 **	0,223 *	0,215 *	0,344 **	0,152
D.1	0,059	-0,131	0,198 *	0,131	0,050	-0,004	0,519 **
D.1 s1	0,051	0,220 *	0,277 **	0,574 **	0,087	0,188 *	0,384 **
D.2	0,023	0,031	0,353 **	0,073	0,017	0,044	0,387 **
D.3	-0,207 *	-0,062	-0,027	0,021	-0,153	-0,144	0,528 **
D.4	-0,133	0,006	0,042	0,006	-0,161	-0,147	0,541 **
D.5	0,108	-0,051	0,356 **	0,189 *	-0,002	0,040	0,512 **
D.6	0,019	0,038	0,205 *	-0,051	-0,035	0,050	0,436 **
D.7	-0,057	0,068	0,204 *	0,135	-0,059	0,034	0,644 **
D.8	0,135	0,046	0,483 **	0,287 **	0,117	0,263 **	0,488 **
D.9	0,003	0,057	0,220 *	0,230 *	0,108	0,050	0,608 **

BSABS-Item-Nr.	Cluster						
	BS1	BS2	BS3	BS4	BS5	BS6	BS7
D.10	a	a	a	a	a	a	a
D.11	-0,136	-0,126	0,368 **	0,170	-0,123	-0,022	0,675 **
D.14	0,009	0,004	0,200 *	0,306 **	-0,004	-0,001	0,618 **

** Die Korrelation ist auf dem 1%-Niveau bei zweiseitiger Testung signifikant.

* Die Korrelation ist auf dem 5%-Niveau bei zweiseitiger Testung signifikant.

^a Kann nicht berechnet werden, da mindestens eine der Variablen konstant ist.

Hellgrau unterlegt Korrelation von Einzelvariable mit ihrem zugeordneten Cluster.

Fettdruck Höchste Korrelation von einer Einzelvariable mit einem Cluster, das nicht das eigene ist.

Für die Numerierungen der BSABS-Items und Cluster siehe Tabelle 18, Kapitel 4.1.2.

ANHANG I: Diagnostische Gütekriterien der Basissymptome in der Berechnungsstichprobe
(n = 120)

Symptom	Sensitivität	Spezifität	PPP	NPP	DLR ⁺	DLR ⁻	%F-P	%F-N
A.5	0,18	0,74	0,37	0,52	0,69	1,11	14,2	37,5
A.6.1	0,64	0,35	0,46	0,54	0,98	1,03	35,0	16,7
A.6.2	0,13	0,94	0,64	0,56	2,17	0,93	3,3	40,0
A.6.3	0,49	0,51	0,46	0,54	1,00	1,00	26,7	23,3
A.6.4	0,35	0,69	0,49	0,56	1,13	0,94	16,7	30,0
A.7.1	0,24	0,71	0,41	0,52	0,83	1,07	15,8	35,0
A.7.2	0,22	0,86	0,57	0,57	1,57	0,91	13,8	35,8
A.8.1 + B.1.2	0,38	0,77	0,58	0,59	1,65	0,81	12,5	28,3
A.8.2 + B.1.3	0,42	0,59	0,46	0,54	1,02	0,98	22,5	26,7
A.8.3 + B.1.4	0,27	0,68	0,42	0,52	0,84	1,07	17,5	33,3
A.8.4	0,26	0,94	0,78	0,60	4,33	0,79	3,3	34,2
B.1.1	0,40	0,60	0,46	0,54	1,00	1,00	21,7	27,5
B.2.1	0,11	0,88	0,43	0,54	0,92	1,01	6,7	40,8
B.2.2	0,27	0,72	0,46	0,54	0,96	1,01	15,0	33,3
B.2.3	0,11	0,89	0,46	0,54	1,00	1,00	5,8	40,8
B.3.1	0,29	0,75	0,50	0,54	1,16	0,95	13,3	32,5
B.3.4	0,18	0,86	0,53	0,55	1,29	0,95	7,5	37,5
C.1.1	0,44	0,91	0,80	0,66	4,89	0,62	5,0	25,8
C.1.2	0,35	0,88	0,71	0,61	2,92	0,74	6,7	30,0
C.1.3	0,38	0,95	0,88	0,65	7,60	0,65	2,5	28,3
C.1.4	0,36	0,86	0,69	0,62	2,57	0,74	7,5	29,2
C.1.5	0,78	0,34	0,50	0,65	1,18	0,65	35,8	10,0
C.1.6	0,40	0,94	0,85	0,65	6,67	0,64	3,3	27,5
C.1.7	0,24	0,92	0,72	0,59	3,00	0,83	4,2	35,0
C.1.8	0,15	0,85	0,44	0,54	1,00	1,00	8,3	39,2
C.1.9	0,22	0,82	0,50	0,55	1,22	0,95	10,0	35,8
C.1.10	0,06	0,94	0,43	0,54	1,00	1,00	3,3	43,3
C.1.12	0,42	0,71	0,55	0,59	1,45	0,82	15,8	26,7
C.1.13	0,26	0,75	0,47	0,54	1,04	0,99	13,3	34,2
C.1.15	0,24	0,95	0,81	0,60	4,80	0,80	2,5	35,0
C.1.16	0,06	10,00	10,00	0,56	E	0,94	0	43,3
C.1.17	0,38	0,91	0,78	0,63	4,22	0,68	5,0	28,3
C.2.1 s1	0,15	0,92	0,62	0,56	1,88	0,92	4,2	39,2
C.2.1 s3	0,06	10,00	10,00	0,56	E	0,94	0	43,3

Symptom	Sensitivität	Spezifität	PPP	NPP	DLR ⁺	DLR ⁻	%F-P	%F-N
C.2.2 s1	0,22	0,97	0,86	0,59	7,33	0,80	1,7	35,8
C.2.2 s2	0,16	0,97	0,82	0,58	5,33	0,87	1,7	38,3
C.2.3 s1	0,04	0,99	0,67	0,55	4,00	0,97	0,8	44,2
C.2.3 s2	0,06	0,99	0,75	0,55	6,00	0,95	0,8	43,3
C.2.3 s3	0,02	10,00	10,00	0,55	E	0,98	0	45,0
C.2.3 s4	0,09	0,97	0,71	0,56	3,00	0,94	1,7	41,7
C.2.3 s5	0,13	10,00	10,00	0,58	E	0,87	0	40,0
C.2.3 s6	0,22	0,95	0,80	0,59	4,40	0,82	2,5	35,8
C.2.3 s7	0,06	0,99	0,75	0,55	6,00	0,95	0,8	43,3
C.2.3 s8	0,02	10,00	10,00	0,55	E	0,98	0	45,0
C.2.3 s9	0,04	0,99	0,67	0,55	4,00	0,97	0,8	44,2
C.2.3 s10	0,00	10,00	0,00	0,54	E	1,00	0	45,8
C.2.3 s12	0,04	0,99	0,67	0,55	4,00	0,97	0,8	44,2
C.2.4 s1	0,27	0,80	0,54	0,57	1,35	0,91	10,8	33,3
C.2.4 s2	0,15	0,92	0,62	0,56	1,88	0,92	4,2	39,2
C.2.5 s1	0,20	0,92	0,69	0,58	2,50	0,87	4,2	36,7
C.2.5 s2	0,02	10,00	10,00	0,55	E	0,98	0	45,4
C.2.6	0,24	0,94	0,77	0,59	4,00	0,81	3,3	35,0
C.2.8	0,20	0,99	0,92	0,59	20,00	0,81	0,8	36,7
C.2.9	0,13	10,00	10,00	0,58	E	0,87	0	40,0
C.2.10	0,06	0,97	0,60	0,55	2,00	0,97	1,7	43,3
C.2.11	0,26	0,89	0,67	0,59	2,36	0,83	5,8	34,2
C.3.1	0,11	10,00	10,00	0,57	E	0,89	0	40,8
C.3.2	0,16	0,94	0,69	0,57	2,67	0,89	3,3	38,3
C.3.3	0,22	0,95	0,80	0,59	4,40	0,82	2,5	35,8
D.1	0,15	0,82	0,40	0,53	0,83	1,04	10,0	39,2
D.1 s1	0,18	0,86	0,53	0,55	1,29	0,95	7,5	37,5
D.2	0,11	0,86	0,40	0,53	0,79	1,03	7,5	40,8
D.3	0,19	0,68	0,32	0,49	0,59	1,19	17,5	37,5
D.4	0,07	0,83	0,27	0,51	0,41	1,19	9,2	42,5
D.5	0,06	0,91	0,33	0,53	0,67	1,03	5,0	43,3
D.6	0,18	0,75	0,39	0,52	0,72	1,09	13,3	37,5
D.7	0,47	0,68	0,55	0,60	1,47	0,78	17,5	24,2
D.8	0,24	0,88	0,62	0,58	2,00	0,86	6,7	35,0
D.9	0,22	0,79	0,46	0,54	1,05	0,99	11,7	35,8
D.10	0,00	10,00	0,00	0,54	E	1,00	0	45,8

Symptom	Sensitivität	Spezifität	PPP	NPP	DLR ⁺	DLR ⁻	%F-P	%F-N
D.11	0,16	0,82	0,43	0,54	0,89	1,02	10,0	38,3
D.14	0,27	0,69	0,43	0,53	0,87	1,06	16,7	33,3

PPP: Positive Prädiktive Stärke (Power)

NPP: Negative Prädiktive Stärke (Power)

DLR⁺: Positive Diagnostische Likelihood Ratio

DLR⁻: Negative Diagnostische Likelihood Ratio

%F-P: Prozentsatz Falsch-Positiver Vorhersagen

%F-N: Prozentsatz Falsch-Negativer Vorhersagen

E: Nicht zu berechnen, da unerlaubte Division durch 0

ANHANG J: Übersicht über die in die einzelnen Vorhersagemodelle aufgenommenen Variablen.

BSABS-Item-Nr.	DA-I	DA-C	LR-I _F	LR-I _B	LR-C	GMM	SPAN	DEI	DLR
A.8.4		X			X			X	X
C.1.1		X			X	X	X	X	X
C.1.2	X			X		X	X	X	
C.1.3	X	X	X	X	X	X	X	X	X
C.1.4		X			X	X		X	
C.1.6	X	X	X	X	X	X	X	X	X
C.1.7		X			X				
C.1.15									X
C.1.16		X			X				
C.1.17		X	X		X	X		X	X
C.2.2 s1			X						X
C.2.2 s2									X
C.2.3 s1									X
C.2.3 s2									X
C.2.3 s6	X			X					X
C.2.3 s7									X
C.2.3 s9									X
C.2.3 s12									X
C.2.6									X
C.2.8									X
C.2.9		X			X				
C.2.11								X	
C.3.3									X

DA-I: schrittweise und nonparametrische Diskriminanzanalyse der Einzelitem

DA-C: schrittweise und nonparametrische Diskriminanzanalyse der Cluster

LR-I_F: schrittweise logistische Regressionsanalyse der Einzelitems im Vorwärtssuchmodus (ausgewählt)

LR-I_B: schrittweise logistische Regressionsanalyse der Einzelitems im Rückwärtssuchmodus (nicht ausgewählt)

LR-C: schrittweise logistische Regressionsanalyse der Cluster

GMM: Grob-Mengen-Modell-basierte Analyse mit DataLogic

SPAN: Search Partition Analysis

DEI: Selektion nach diagnostischen Effizienzindizes

DLR: Selektion nach positiven diagnostischen Likelihood Ratios

Für die Bezeichnungen der Basissymptome siehe Anhang C.

ANHANG K: Übersicht über die Gruppenzuordnungen der einzelnen Vorhersagemodelle für die Probanden der Validierungsstichprobe (n = 40).

Patienten Nr.	beob. GZ	vorhergesagte Gruppenzugehörigkeit (GZ)								falsch klassi.
		DA-I	DA-C	LR-I _F	LR-C	GMM	SPAN	DEI	DLR	
3	1	0	0	0	0	0	0	0	0	8
16	0	0	0	0	0	0	0	0	0	0
24	1	1	1	1	0	1	1	0	1	2
30	1	1	1	0	1	1	1	1	1	1
31	1	1	1	1	1	1	1	1	1	0
32	1	1	1	1	1	1	1	1	1	0
35	1	1	1	1	1	1	1	1	1	0
37	0	1	1	1	1	1	1	1	0	7
44	1	1	1	1	1	1	1	1	1	0
48	1	0	0	0	0	0	0	0	0	8
51	1	1	1	1	1	1	1	1	0	1
53	1	0	0	0	0	0	0	0	0	8
54	0	1	1	1	0	1	1	0	0	5
55	0	0	0	0	0	0	0	0	0	0
62	0	1	1	1	1	1	1	1	1	8
64	1	1	1	1	1	1	1	1	1	0
65	0	0	1	1	0	0	0	1	0	3
68	1	1	1	1	0	1	1	1	1	1
74	1	1	1	1	1	0	1	1	1	1
75	0	1	0	0	0	0	1	0	0	2
82	1	0	1	1	1	1	1	1	1	1
90	0	0	0	0	0	0	0	0	0	0
97	0	1	1	1	0	1	1	1	0	6
103	0	0	0	0	0	0	0	0	0	0
105	1	0	0	0	0	0	0	0	0	8
106	1	1	1	1	1	0	1	1	1	1
111	1	1	1	1	1	1	1	1	1	0
113	1	1	1	1	1	1	1	1	1	0
120	1	1	1	1	1	0	1	0	0	3
122	0	0	0	0	0	0	0	0	0	0
127	0	0	1	1	0	0	0	0	0	2
128	1	1	1	1	1	1	1	1	1	0
131	1	0	1	1	1	1	1	1	1	1

Patienten Nr.	beob. GZ	vorhergesagte Gruppenzugehörigkeit (GZ)								falsch klassi.
		DA-I	DA-C	LR-I _F	LR-C	GMM	SPAN	DEI	DLR	
133	1	0	0	0	0	0	0	0	0	8
140	0	0	1	0	0	0	0	0	0	1
143	0	0	0	0	0	0	0	0	0	0
151	0	0	1	0	0	0	0	1	0	2
154	1	1	1	1	1	1	1	1	1	0
155	0	0	0	0	0	0	0	0	0	0
160	1	1	1	1	1	1	1	1	0	1

beob. GZ: beobachtete Gruppenzugehörigkeit

0: Zugehörigkeit zu Gruppe ohne Schizophrenieentwicklung

1: Zugehörigkeit zu Gruppe mit Schizophrenieentwicklung

falsch klassi.: Anzahl fälschlicher Klassifizierungen des Patienten durch die Vorhersagemodelle

DA-I: schrittweise und nonparametrische Diskriminanzanalyse der Einzelitems

DA-C: schrittweise und nonparametrische Diskriminanzanalyse der Cluster

LR-I_F: schrittweise logistische Regressionsanalyse der Einzelitems im Vorwärtssuchmodus

LR-C: schrittweise logistische Regressionsanalyse der Cluster

GMM: Grob-Mengen-Modell-basierte Analyse mit DataLogic

SPAN: Search Partition Analysis

DEI: Selektion nach diagnostischen Effizienzindizes

DLR: Selektion nach positiven diagnostischen Likelihood Ratios

ANHANG L: Lebenslauf

geboren: 08.02.1966 in Göttingen

Staatsangehörigkeit: deutsch

Schulische Ausbildung

1972 Einschulung in die Mittelpunktschule in Dassel
1976 Wechsel auf die Paul-Gerhardt-Schule in Dassel
1985 Erlangung der Hochschulreife

Berufliche Ausbildung

1986-1993 Studium im Diplom-Studiengang 'Psychologie' an der Biologischen Fakultät der Georg-August-Universität in Göttingen mit den Schwerpunktbereichen Kognitionspsychologie und Entwicklungspsychologie.
1988-1990 Besuch von Vorlesungen im Bereich Publizistik
1993 Diplomarbeit im Fach Psychologie: *Basissymptome bei geistig Behinderten mit Schizophrenie - Eine Vergleichsstudie mit Schizophrenen und geistig Behinderten.*

Beruflicher Werdegang

01.06.1994 - 31.07.1996 wissenschaftliche Mitarbeiterin an der Psychiatrischen Klinik der Rheinisch-Westfälischen Technischen Hochschule, Aachen, Schwerpunkt 'Früherkennung von Schizophrenie', davon ein Jahr als Projektleiterin des DFG-Projekts KI 970/1-1, 'Früherkennung der Schizophrenie'.
01.08.1996 – 30.06.2000 wissenschaftliche Mitarbeiterin an der Klinik für Psychiatrie und Psychotherapie der Universität zu Köln, Schwerpunkt 'Früherkennung von Schizophrenie' und Aufbau des Früherkennungs- und -therapiezentrams für psychotische Krisen, FETZ.
seit 01.07.2000 wissenschaftliche Mitarbeiterin am Zentralinstitut für Seelische Gesundheit, Mannheim, Arbeitsgruppe Schizophrenieforschung, Schwerpunkt: methodisch-statistische Mitarbeit im

Teilprojekt 1.1.1., ‚Früherkennung der Schizophrenie‘, des BMFT-geförderten Kompetenznetzwerk Schizophrenie. Weitere Aufgaben: Mitarbeit an der Deutsch-Israelischen Projektkooperation DIP-C 1.2, ‚Early Detection and Intervention in Schizophrenia‘; Projektkoordination des Kölner DFG-Projekts KI 970/3-1, ‚Prädiktion der Schizophrenie‘; methodisch-statistische Betreuung des Stanley Research Awards ‚Evaluation of the antipsychotic efficacy of the cannabinoid-antagonistic cannabidiol in treating acute schizophrenic psychosis. A double-blind, controlled clinical trial‘.

Sonstiges

01.08.1985 - 31.12.1985	Au-pair-Aufenthalt in Bath / England
1993	Erwerb des Cambridge Certificate of Proficiency