

Endbericht zum Forschungsvorhaben

# **Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung in der Psychiatrie**

Bericht an den Bundesminister für Gesundheit

## **Autoren:**

Ulrich Frick  
Stefan Krischker  
Clemens Cording

Mit einem Kapitel von Winfried Barta und Harald Binder



**AG Versorgungsforschung**  
Regensburg

## **Anschrift:**

Psychiatrische Universitätsklinik Regensburg  
Arbeitsgruppe Versorgungsforschung  
Prof. Dr. C. Cording  
Universitätsstraße 84  
93042 Regensburg

## Gliederung

Abschnitt	Seite
0. Vorwort	3
1. Datenqualität der BADO	5
1.1 Vollständigkeit	6
1.2 Zuverlässigkeit (Reliabilität) der Daten	7
1.3 Inhaltliche Gültigkeit (Validität) der BADO-Daten	10
2. Statistische Methoden zum Krankenhausvergleich	14
2.1 Ausgangsidee	14
2.2 Level-1-Prädiktoren: Einfluss von Patientenmerkmalen	17
2.3 Level-2-Prädiktoren: Einfluss von Organisationsmerkmalen	18
3. Stichprobe eines beispielhaften Krankenhausvergleiches	22
3.1 Datenabgleich	22
3.1.1 Datengewinnung	23
3.1.2 Datenaufbereitung und –prüfung	24
3.1.3 Verbesserte Datenerhebung in den Kliniken	25
3.1.4 Strukturanalyse	26
3.1.5 Datenschutz	26
3.2 Deskriptiver Vergleich ohne statistische Adjustierung	28
3.2.1 Demographische und anamnestische Daten der PatientInnen	28
3.2.2 Merkmale des Behandlungsprozesses in den Kliniken	35
3.2.3 Merkmale des Behandlungsergebnisses	39
4. Ergebnisse des Multi-Level-Vergleichs von acht Kliniken	42
4.1 "Where have all those differences gone?"	42
4.2 Einflussfaktoren auf die Verweildauer	45
4.2.1 Über die Stationen konstante Patienten-Effekte	47
4.2.2 Über die Stationen variierende Effekte	48
4.2.3 Stationseffekte ohne Wechselwirkung mit Patienten-Merkmalen	52
4.2.4 Stationseffekte in Wechselwirkung mit Patienten-Merkmalen („cross-level interaction“)	53
4.2.5 Kurzzusammenfassung und Tabellierung der Parameter	54
4.3 Diskussion der Stationseinflüsse auf die Verweildauer	57
4.4 Methodisches Fazit	64
5. Vergleiche der Effektivität und Kosten-Effektivität (Autoren: Winfried Barta und Harald Binder)	65
5.1 Einleitung	65
5.2 Fragestellung	65
5.3 Ergebnisse	67
5.4 Diskussion	73
6. Konsequenzen und Ausblick	75
7. Literatur	79

## 0. Vorwort

Krankenhausvergleiche zum Zweck der externen Qualitätssicherung werden seit langem gesetzlich gefordert (§§ 137 bzw. 137a, 137b SGB V; § 5 BpflV), sind aber bisher allenfalls ansatzweise zur Routinepraxis geworden. Je nach medizinischem Fachgebiet ergeben sich unterschiedliche Problemstellungen. Psychiatriespezifische Probleme sind die besondere Komplexität und Mehrdimensionalität der Störungsbilder, der Behandlungsziele, der dazu eingesetzten Verfahren sowie der zu beurteilenden Behandlungsergebnisse. Anders als beispielsweise in den operativen Fächern sind die psychiatrischen Behandlungsverläufe viel stärker von individuellen Besonderheiten mitgeprägt. Gleichwohl hat die Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) relativ frühzeitig speziell im Hinblick auf die Qualitätssicherung eine neue, bundesweit gültige psychiatrische Basisdokumentation (BADO) für den (teil-)stationären Bereich entwickelt (Cording et al., 1995), die auch von der Bundesdirektorenkonferenz, von der Konferenz der psychiatrischen Lehrstuhlinhaber, von der Aktion Psychisch Kranke, von der Bundesarbeitsgemeinschaft der Träger psychiatrischer Krankenhäuser sowie von den Spitzenverbänden der Krankenkassen anerkannt und empfohlen wird. Die DGPPN-BADO wurde von Anfang an nicht nur für Zwecke der internen, sondern auch für die *externe* Qualitätssicherung durch *Krankenhausvergleiche* entwickelt (vgl. Cording 1997).

Mit dem durch das Bundesministerium für Gesundheit geförderten Projekt „Freiwillige Krankenhausvergleiche mit der DGPPN-BADO zur externen Qualitätssicherung in der Psychiatrie“ sollten Verfahren entwickelt und erprobt werden, wie die gesetzlichen Vorgaben im Bereich der (teil-)stationären Psychiatrie mit Hilfe des Merkmalskataloges der DGPPN-BADO so umgesetzt werden können, dass das Ziel der Qualitätssicherung auf der Basis geeigneter Daten und angemessener methodischer Verfahren erreicht wird.

Im Unterschied zu anderen Projekten, die sich dabei (wie in manchen medizinischen Fachgebieten üblich) auf eine Subgruppe sog. Tracer-Diagnosen beschränken, gehen bei unserem Modell alle Patientengruppen in den Krankenhausvergleich ein, weil bei dem weitgefächerten Spektrum psychiatrischer Störungsbilder die Generalisierbarkeit der Qualität beispielsweise der Suchtbehandlung auf die Qualität der Klinik insgesamt nicht sicher vorausgesetzt werden kann.

Bewusst wurde darauf verzichtet, über die für die Routinedokumentation vorgesehene, klinisch multifunktional nutzbare DGPPN-BADO hinaus spezielle Erhebungsinstrumente eigens für den Zweck der Krankenhausvergleiche bzw. der externen Qualitätssicherung zusätzlich einzuführen.

Unerwartete Aktualität haben die Ergebnisse unseres Projekts dadurch gewonnen, dass die Krankenhausvergleiche nach § 5 BpflV im Zuge der Einführung der Diagnoses Related Groups (DRGs) in den somatischen Fächern an Bedeutung verlieren werden, während sie im Bereich der (teil-)stationären Psychiatrie und Psychotherapie nun offenbar mit umso größerer Intensität vorangetrieben werden sollen.

Unser Projekt hat zu wichtigen methodischen Erkenntnissen geführt, die bei der praktischen Umsetzung dieser Krankenhausvergleiche unbedingt berücksichtigt werden sollten, um gravierende Fehlschlüsse mit unabsehbaren Folgen für die Qualität der psychiatrischen Versorgung zu vermeiden (siehe hierzu auch Cording et al. 2001, sowie das Memorandum der DGPPN zum Krankenhausvergleich nach § 5 BpflV).

Wir danken dem Bundesministerium für Gesundheit für seine finanzielle Förderung, ohne die dieses Projekt nicht durchführbar gewesen wäre.

Prof. Dr. Clemens Cording  
Leiter der Arbeitsgruppe Versorgungsforschung  
Psychiatrische Universitätsklinik am Bezirksklinikum Regensburg

## **1. Datenqualität der BADO**

Die Eignung des von der zuständigen Fachgesellschaft (DGPPN) empfohlenen Erhebungsinstruments BADO für die Basisdokumentation in der stationären Psychiatrie (Cording et al., 1995) wurde im vorliegenden Forschungsprojekt zu Beginn der Arbeiten einer breiten Überprüfung hinsichtlich seiner Tauglichkeit für externe, qualitätssichernde Krankenhausvergleiche unterzogen.

Dazu diente folgendes Prozedere:

- 1) Überprüfung der mit der DGPPN-BADO erzielbaren Datenqualität nach psychometrischen Gütekriterien
- 2) Aufbau einer gemeinsamen Datenbank pseudonymisierter Behandlungsdatensätze von Behandlungsepisoden aus acht zum Vergleich bereiten Kliniken
- 3) Erprobung geeigneter statistischer Methoden zum "fairen Vergleich" von verschiedenen Krankenhäusern anhand der gemeinsamen Datenbank
- 4) Analyse der Daten.

Im Nachfolgenden werden zu diesen 4 Schritten die wichtigsten Arbeitsprozesse und ein kurzer Abriss der Ergebnisse dargestellt.

Die (gekürzten) Ergebnisse dieses ersten Abschnittes im Forschungsvorhaben wurden auch breiter dargestellt bei Frick et al. (1999): Die hier berichteten Analysen wurden vorgenommen anhand der gepoolten BADO-Datensätze aus den Kliniken Mainkofen, Regensburg und Landshut (Aufnahmen des Jahres 1996, die im selben Jahr auch wieder entlassen wurden; N=8874). Zwischen diesen drei im ostbayerischen Raum benachbarten Kliniken bestand bereits im Vorfeld dieser Studie eine fruchtbare Kooperation bei der Auswertung der BADO-Daten. In allen drei Kliniken war die BADO zum Untersuchungszeitpunkt bereits im Routineeinsatz, so dass von realistischen Zahlen bei der Vollständigkeitsüberprüfung auszugehen ist. Die Stichprobe überschneidet sich nicht mit den Datensätzen in der Datenbank des Arbeitsschrittes 3 (statistischer Vergleich). Dort wurden in einem anderen Zeitfenster zusätzlich zu den drei genannten Kliniken 5 weitere Häuser zur gemeinsamen Auswertung gewonnen.

Für Krankenhausvergleiche sind verlässliche Daten eine notwendige Voraussetzung. Geprüft wurden daher in diesem ersten Schritt

1. die erzielte Vollständigkeit der Dokumentation auf Ebene der einzelnen Variablen pro Behandlungsebene,
2. die Reliabilität zentraler diagnostischer Aussagen der DGPPN-BADO, sowie
3. Aspekte der inhaltlichen Validität der erhobenen Variablen.

### **1.1 Vollständigkeit**

Im Routinebetrieb wurden von 45 überprüften Dokumentationsvariablen (Zufallsauswahl der Variablen) insgesamt 35 mit Raten von deutlich über 90% vollständig aus-

gefüllt. Variablen mit höheren Fehlquoten berühren vor allem anamnestische Sachverhalte, die insbesondere bei verwirrten PatientInnen zum Aufnahmezeitpunkt nicht erfragbar sind. Kumuliert man pro Behandlungsfall die Anzahl der fehlenden Werte, so bleiben über 80% der PatientInnen unterhalb einer "Fehlende-Werte-Summe" von 4 Variablen (aus 45 überprüften Merkmalen).

Es lässt sich also im Routinebetrieb eine hohe Dokumentationsdichte festhalten. Weil jedoch der Anspruch einer vollständigen Basisdokumentation immer nahe an einer 100%igen Ausschöpfungsquote der Falldokumentation zu liegen hat, wurden die aufgetretenen Missing Values einer detaillierten Analyse unterzogen. Ziel dieser Überprüfung war es, zu klären, ob Missing Values unsystematisch über alle PatientInnen streuen und daher durch eine allgemeine Erhöhung der Dokumentations-Compliance der behandelnden Ärzte zu vermindern sind, oder ob typische Muster von fehlenden Werten bei bestimmten PatientInnengruppen auftreten. Im letzteren Falle müssten zur Erhöhung der Ausschöpfungsquote spezifische Maßnahmen vorgeschlagen werden.

Typische Muster von fehlenden Werten wurden mithilfe der Latenten-Klassen-Analyse voneinander isoliert (vgl. Frick et al., 1999). Die Gründe für aufgetretene fehlende Werte konnten gut den Mustern an fehlenden Variablen zugeordnet werden:

- Klasse 1 "Weitgehend vollständige Bögen" (rund 80%) - keine Verbesserung notwendig
- Klasse 3 "Weitgehend fehlende Bögen" (rund 6,4%) - keine Verbesserung notwendig, da fast ausschließlich PatientInnen des Regensburger Schlaflabors, ohne direkte Verbindung zu den PatientInnen der Akut-Psychiatrie.
- Klasse 7 "Wechsel von Tagesklinik in vollstationären Aufenthalt" (1,1%) - typischerweise Fehlen von anamnestischen Daten, da diese in der Handakte der Tagesklinik bereits enthalten sind, aber nicht auf EDV-gerechte Bögen übertragen wurden. Verbesserungsmöglichkeit: Optimierung der Verwaltungsabläufe, Einführung von elektronischer Patientendokumentation auch in der Tagesklinik
- Klasse 5 "Schlaflabor" (1,6%) - Typischerweise fehlen hier nur diejenigen Variablen, die eine Beurteilung des Schweregrades der Psychopathologie (CGI-Rating) und der Sozialen Funktionserfüllung (Global Assessment of Functioning GAF) erfordern. Dieses Rating ist für diese PatientInnen-Subgruppe wenig zielführend.
- Klasse 2 "Drehtür-PatientInnen" (8,9%) - Vor allem anamnestische Angaben (Schulabschluss, Ersthospitalisierung, berufliche Situation, Familienstand etc.) sind die hier fehlenden Variablen. Sie entstehen vor allem deshalb, weil die hier vertretene PatientInnengruppe extrem kurze Verweildauern und die Diagnose "Alkoholabhängigkeit/-missbrauch" aufweisen, und von daher eine Erhebung der Angaben oft nicht möglich ist. Die fehlenden Werte lassen sich jedoch in einer Vielzahl der Fälle aus früheren Angaben zu einem früheren Aufnahmezeitpunkt ergänzen. Mit Einführung einer elektronischen PatientInnen-Akte würden sie völlig wegfallen.
- Klasse 6 "Fehlende Aufnahmebögen bei Sucht / Erstaufnahme" (1,3%) - Auch hier fehlen zumeist Anamnesedaten bei PatientInnen mit Alkoholsuchtproblemen. Vermutlich sind Schwierigkeiten bei der anamnestischen Exploration für diesen Typ von Missing Values verantwortlich.
- Klasse 4 "Betreuungsrecht / Fixierungen" (1,2%) - Für diese PatientInnen-Gruppe fehlen zumeist die Einschätzungen der sozialen Anpassung im Vorfeld

der Hospitalisierung. Die weiteren Prozessvariablen sind zumeist vollständig. Es scheint so, dass bei dieser Patientengruppe diese Daten schwer beschaffbar sind, da Krankheitseinsicht, gute Gedächtnisleistungen und Kooperationsbereitschaft oft nicht gegeben sind.

### **Gesamtwertung der Missing-Value-Analyse:**

Ein Gutteil der fehlenden Werte sind völlig unproblematisch, weil einesteils bei PatientInnen außerhalb der Akutpsychiatrie entstanden, oder weil sie gut aus konsekutiven Aufnahmen substituierbar sind. Zwei organisatorische Veränderungen im Dokumentationsprozess wurden zur Verbesserung bestehender Schwachstellen angeregt: Verbesserung der Datenübergabe bei Einweisung aus hauseigenen Tageskliniken, und Vorbereitung zur Entlassung auch formal unter Prüfung der bislang erreichten Vollständigkeit in der Dokumentation.

## **1.2 Zuverlässigkeit (Reliabilität) der Daten**

Reliabilität kann psychometrisch über mehrere Verfahren quantifiziert werden: a) als Inter-Rater-Übereinstimmung, b) als Intra-Rater-Übereinstimmung, c) über die interne Konsistenz eines zusammengesetzten Maßes.

Eine wichtige Kernvariable der BADO ist natürlich die vom behandelnden Arzt gestellte **Entlassungsdiagnose** eines Patienten. Der Königsweg zur Bestimmung der Zuverlässigkeit dieser Diagnoserhebung in der BADO besteht darin, die von den ÄrztInnen routinemäßig erhobenen Diagnosen mit denjenigen Diagnosen zu vergleichen, die aufgrund eines standardisierten psychiatrischen Interviews mit bekannten psychometrischen Eigenschaften entstanden sind. Als solche "gold standards" des Diagnosenvergleichs können das WHO Composite International Diagnostik Interview (CIDI) oder das Strukturierte Klinische Interview für DSM-IV (SKID vgl. Wittchen et al., 1997) dienen.

Im Rahmen der BMG-Studie wurde für den Zweck der Diagnosen-Validierung der SKID durch 2 geschulte Interviewer an n=108 zufällig ausgewählten PatientInnen des Bezirkskrankenhauses Regensburg durchgeführt. Diese Fallzahl wurde als erforderliche Stichprobengröße berechnet, ab der ein mögliches Unterschreiten einer vorab festgelegten Mindestgüte von Cohen's kappa = 0.7 (vgl. Cohen, 1960) um 0.2 Punkte auf kappa = 0.5 mit einer als ausreichend erachteten statistischen Power von 80% festgestellt werden kann (Berechnung mittels nQuery 3.0). Die Interviews konnten bis zum Oktober 2000 vollständig durchgeführt werden. Wegen einer ernsthaften Erkrankung des involvierten Doktoranden jedoch konnten die erhobenen SKID-Daten erst mit einer großen zeitlichen Verzögerung mit den routinemäßig an denselben 100 PatientInnen auf dem üblichen Weg erhobenen Daten der BADO zusammengeführt werden. Die statistischen Analysen sind derzeit in Arbeit und werden nun gesondert veröffentlicht.

Parallel zum Projekt der Diagnosenvalidierung via externe, standardisierte psychiatrische Interviews wurde ein zweiter Ansatz verfolgt, der ebenfalls die Verlässlichkeit der Diagnosenstellung im Rahmen der BADO-Dokumentation abschätzen sollte. Ausgangsidee war die Überlegung, dass bestimmte chronische Krankheiten bei Wiederaufnahmen derselben Patienten im Klinikum zur identischen Diagnosenstellung führen sollten.

**Tab.3 Diagnosenübereinstimmung bei Mehrfachaufnahmen in 1996**

Diagnose (N bei Erstaufnahme)	dieselbe Diagnose bei Rehospitalisierungen?					
	1. Aufnahme versus:			2. Aufnahme versus:		3. Aufnahme versus:
Cohen's Kappa McNemar Test	2. Aufnahme n=1270	3. Aufnahme n=389	4. Aufnahme n=150	3. Aufnahme n=393	4. Aufnahme n=150	4. Aufnahme n = 149
<b>Schizophrenie (233)</b>	0,831 ./.	0,819 ./.	0,702 ./.	0,849 ./.	0,702 ./.	0,764 ./.
<b>schizo-affektive Psychose (47)</b>	0,742 ./.	0,704 ./.	0,906 ./.	0,821 ./.	0,656 ./.	0,885 ./.
<b>affektive Psychose (102)</b>	0,785 ./.	0,707 ./.	0,885 ./.	0,808 ./.	0,701 ./.	0,827 ./.
<b>sonstige endogene Psychose (8)</b>	0,614 ./.	not appl.	n.a.	n.a.	n.a.	n.a.
<b>Oligophrenie (32)</b>	0,836 ./.	0,838 ./.	0,797 ./.	0,795 ./.	1 ./.	1 ./.
<b>organische Psychose (120)</b>	0,828 ./.	0,799 ./.	0,656 ./.	0,911 ./.	0,744 ./.	0,885 ./.
<b>nicht-psychotische hirnorganische Erkrankg. (10)</b>	0,531 ./.	n.a.	n.a.	n.a.	n.a.	n.a.
<b>Alkoholismus (362)</b>	0,882 ./.	0,866 ./.	0,852 ./.	0,889 ./.	0,824 ./.	0,864 ./.
<b>Medikamenten-, Drogenabhängigkeit (148)</b>	0,837 ./.	0,803 ./.	0,82 ./.	0,852 ./.	0,748 ./.	0,79 ./.
<b>Neurosen (116)</b>	0,615 ./.	0,293 ./.	0,098 p = 0,092 a)	0,432 ./.	0,237 p = 0,065 a)	0,153 ./.
<b>Persönlichkeitsstörungen (61)</b>	0,68 ./.	0,567 ./.	0,52 ./.	0,655 p = 0,008 b)	0,475 p = 0,021 b)	0,654 ./.
./.	Kappa jeweils bei dichotomisierter Variable (gegen alle anderen Diagnosengruppen)					
a) Abnahme						
b) Zunahme						

Inter-Rater-Übereinstimmung kann daher bei wiederaufgenommenen PatientInnen gemessen werden mittels Kreuztabellierung der zumeist von unabhängigen Ärzten erstellten Entlassungsdiagnosen. Tabelle 3 enthält die dementsprechenden Ergebnisse für insgesamt 11 Diagnosengruppen.

Die Diagnosengruppen -Schizophrenie, -schizo-affektive Psychose, -affektive Psychose, -Oligophrenie und -organische Psychose erzielen dabei durchwegs zufriedenstellende (> 0.7) bis gute (> 0.8), teils sogar sehr gute (> 0.9) Übereinstimmungskoeffizienten. Die Diagnosengruppen "sonstige endogene Psychose" und "nicht-psychotische hirnorganische Erkrankung" bleiben von den Fallzahlen her sehr gering (≤ 10) und treten zudem bei Dritt- und Viertaufnahme nicht mehr in Erscheinung. Alkoholismus und Medikamenten- bzw. Drogenabhängigkeit erzielen gute Übereinstimmungskoeffizienten über alle Aufnahmezahlen hinweg.

Geringere Übereinstimmungen treten bei den Diagnosegruppen "Neurosen" und "Persönlichkeitsstörungen" zutage. Von Aufnahmezahl zu Aufnahmezahl sinkt die Übereinstimmung zur Diagnose bei der Erstaufnahme für die Neurosen stark ab, von zunächst  $\kappa=0,615$  bis auf  $\kappa=0.098$ . Ein Kappa-Wert nahe Null bedeutet nicht, dass keinerlei Übereinstimmung mehr zwischen Erst- und Zweitbeurteilung vorhanden wäre, sondern lediglich, dass sich die "Trefferzahlen" zwischen Erst- und Zweitbeurteilung ganz nahe an dem bewegen, was allein aus den Randwahrscheinlichkeiten beider Merkmale unter statistischer Unabhängigkeit erwartbar ist.

Patienten, die im Jahr 1996 mindestens viermal hospitalisiert wurden, haben dann eine im Vergleich zur Erstaufnahme ( $p=0.092$ ) oder Zweitaufnahme ( $p=0.065$ ) verringerte Chance (McNemar Test auf Symmetrie), beim vierten stationären Aufenthalt immer noch als an einer "Neurose" erkrankt diagnostiziert zu werden. Bei derart intensiver Inanspruchnahme psychiatrischer Hilfen bevorzugen die behandelnden Ärzte diagnostische Kategorien, die vermehrt für chronische Verläufe angewendet werden. Mit anderen Worten: Die Erstdiagnose einer "Neurose" hat sich nachträglich als zu korrigierende herausgestellt, und es handelt sich dann beispielsweise um eine unipolare Depression.

Persönlichkeitsstörungen sind im allgemeinen von vorneherein schwieriger diagnostizierbar als z.B. eine manifeste Psychose (Fiedler, 1995). Dies bildet sich auch in den geringeren Übereinstimmungskoeffizienten von Tabelle 3 ab ( $\kappa$  zwischen 0.47 und 0.68). Während jedoch bei den neurotischen Störungen mit steigender Aufnahmezahl eine Abnahme in der ursprünglichen Diagnosenstellung zu verzeichnen war, können bei den 393 Patienten, die 1996 mindestens ein drittes Mal stationär aufgenommen wurden bzw. bei den 150 Patienten, für die ein vierter Aufenthalt registriert wurde, vermehrt Diagnosenwechsel von anderen Kategorien in die Kategorie "Persönlichkeitsstörung" registriert werden. Das bedeutet möglicherweise, dass Persönlichkeitsstörungen, die ja per definitionem als zeitstabil vorausgesetzt werden, erst mit zunehmender Bekanntheit mit dem Patienten und weiterer Kenntnis des Störungsverlaufs sicherer als solche diagnostiziert werden können. Insofern ist die sinkende Übereinstimmung zur "Erstdiagnose" also durchaus als störungsimmanente Besonderheit und weniger als "Makel" einer ungenügenden diagnostischen Erfahrung der BADO-dokumentierenden ÄrztInnen aufzufassen.

Zusammenfassend ergibt sich aus der Prüfung der Inter-Rater-Reliabilität bei Wiederaufnahmen ein zufriedenstellendes Bild der Diagnosenübereinstimmung. Weitere Analysen sind in Vorbereitung. Einschränkend muss hinzugefügt werden, dass die Verlässlichkeit eines Dokumentationssystems zu einem bestimmten Zeitpunkt noch keinerlei Garantie dafür beinhaltet, dass dasselbe Dokumentationssystem zu einem späteren Zeitpunkt oder in einer anderen Klinik dieselbe Reliabilität erzielt. Lassen beispielsweise die Anstrengungen einer Klinik bei der Einschulung neu hinzukommender Ärzte in die BADO schrittweise über die Jahre nach, dann muss damit gerechnet werden, dass sich Messfehler und Fehlkodierungen häufen. Von daher ist für jedes nach der BADO dokumentierende Haus anzuregen, dass intern sogenannte Standard Operating Procedures erstellt werden, die die Wissensweitergabe an die neu eintretenden Ärzte sicherstellen, und die auch im Sinne der Qualitätssicherungsschleifen die Güte der Erhebung in Abständen messen und den ÄrztInnen zurückmelden.

### 1.3 Inhaltliche Gültigkeit (Validität) der BADO-Daten

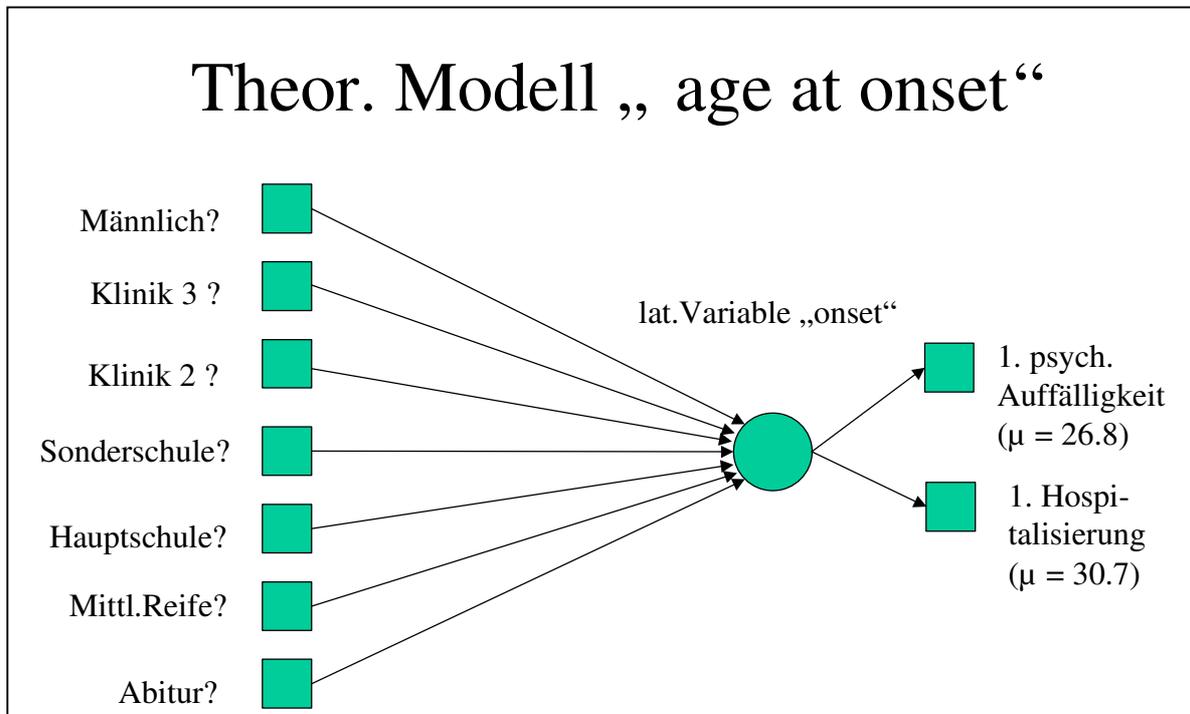
Unter Validität wird in der Psychometrie verstanden, dass ein Messinstrument auch das misst, wozu es konstruiert wurde. Die Messung des Körpergewichts eines Menschen mittels eines Meterbands würde beispielsweise sehr reliable Daten erbringen: von Versuch zu Versuch bestehen beim Messen von Kopf bis Fuß der immer gleichen Person nur ganz geringe Abweichungen. Trotzdem wäre eine Interpretation der Messwerte als „Gewicht“ dieser Person völlig verfehlt, sogar wenn man in Rechnung stellt, dass Körpergröße (was hier gemessen wird) und Körpergewicht eine enge Korrelationsbeziehung haben. Die Bestimmung der Validität eines Messinstruments ist also immer auch eine theoretische Frage. Für die BADO bedeutet das, dass sie immer nur so gut sein kann, wie die psychiatrische Theorie über Krankheitsbilder, Risikofaktoren, ätiologische Faktoren und Verläufe. Im Rahmen dieses Forschungsprojekts haben wir uns daher darauf beschränkt, nur an einem kleinen Beispiel zu prüfen, ob mit einer Routinedokumentation wie der BADO der von der „state-of-the-art-Forschung“ beobachtete Sachstand an einem jüngeren Forschungsthema ebenfalls nachvollzogen werden kann.

Das von uns ausgewählte Beispiel behandelt zwei BADO-Items zum Thema „age at onset“, nämlich die Variablen „Jahr der ersten (teil-)stationären psychiatrischen Behandlung überhaupt“ (A9) und „Alter bei erster psych. Auffälligkeit“ (ZA13). Beide Variablen können als mit Messfehlern behaftete, indirekte Indikatoren eines latenten Konstrukts „age at onset“ aufgefasst werden. Während für A9 vor allem mit Gedächtniseffekten und ungenügenden Ankerreizen beim Erfragen solcher biographischer Daten als möglichen Fehlerquellen zu rechnen ist (Sudman et al., 1996), besteht das Messfehlerproblem für die Variable ZA13 wohl stärker in der begrifflichen Unschärfe von „psych. Auffälligkeit“: psychiatrische? psychische? psychologische? Nur die wenigsten dokumentierenden Ärzte unterziehen sich der Mühe, die Abkürzung im offiziellen BADO-Glossar nachzuschlagen (n.b.: sie bedeutet „psychopathologisch“). Ab wann zählt wer wen unter die „Auffälligen“? Daher haben wir zur Validitätsprüfung dieser beiden Variablen von vorneherein eine explizite Modellierung von Messfehlern in einem geeigneten statistischen Modell (lineares Strukturgleichungsmodell, zur Einführung vgl. Backhaus et al., 1996, Kap. 7) vorgesehen.

Inhaltlich sollten die beiden Indikatoren folgendes leisten: Ein früherer Beginn einer schizophrenen Erkrankung bei Männern wurde in zahlreichen Studien sowohl in Inanspruchnahmepopulationen wie auch in Bevölkerungsstichproben (vgl. Angermeyer & Kuhn, 1988) nachgewiesen. Dabei ist noch nicht endgültig geklärt, ob mit dem vorzufindenden Geschlechtsunterschied unterschiedliche ätiologische Modelle für Schizophrenie bei Männern und Frauen verbunden sind (Östrogen-Hypothese, vgl. Häfner et al., 1993; Lindamer et al., 1997), oder ob der Geschlechtsunterschied größtenteils einen Reflex auf unterschiedlich ausgeprägte Confounder-Variablen (Jablensky & Cole, 1997) darstellt. In jedem Falle sollte der geschlechtsspezifische Unterschied auch in den Indikatoren der BADO auffindbar sein, wenn die Variablen A9 und ZA13 tatsächlich für die Messung des „age at onset“ tauglich sind.

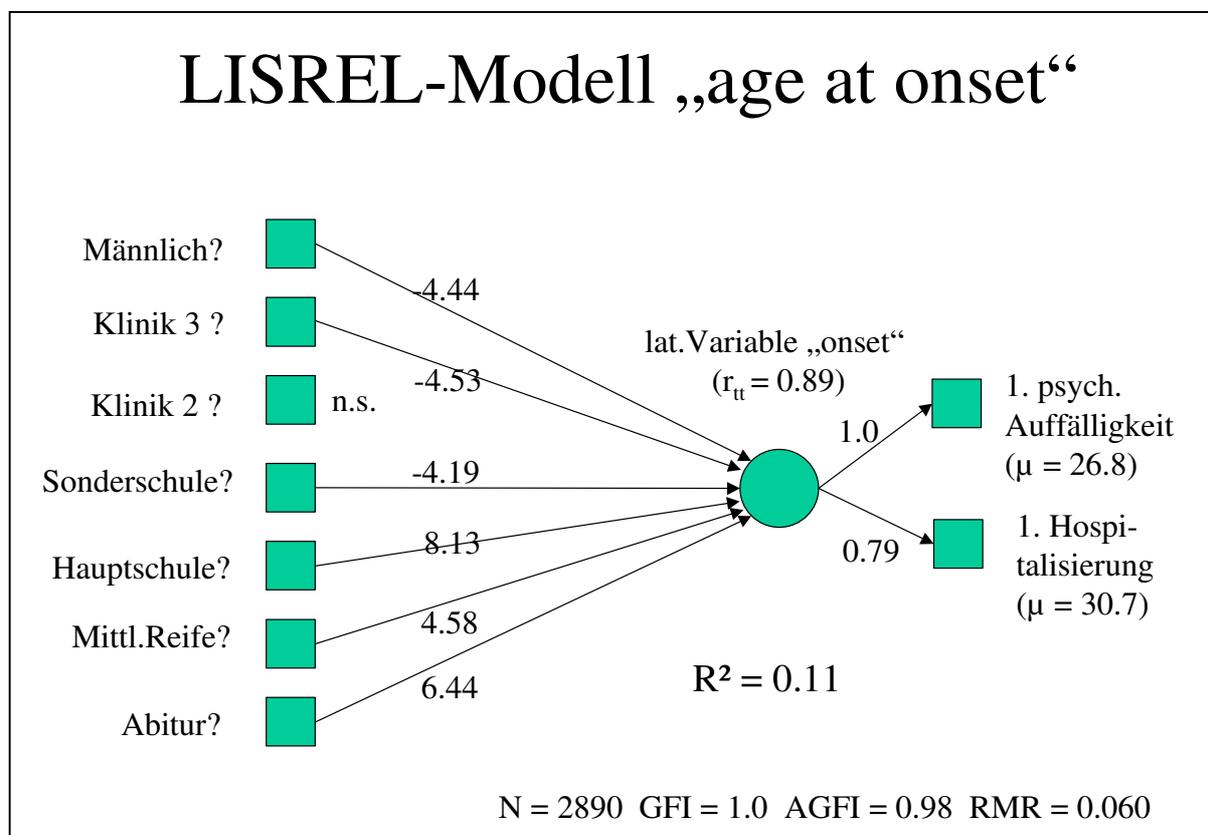
Im unten abgebildeten Modell für die latente Variable „onset“ wurden als Prädiktoren neben dem Geschlecht auch sozioökonomische Bedingungen (via die Variable Bildungsabschluss) als mögliche Confounder berücksichtigt sowie die Klinikzugehörig-

keit als „Schmutzvariable“ für eventuell stattfindende, aber bislang unbekannte Patientenselektionsmechanismen zwischen den drei beteiligten Kliniken. Dabei wurden sowohl der Bildungsabschluss (Referenzkategorie = kein Abschluss) wie die Kliniken (Referenz = Klinik 1) in dichotome Dummies zerlegt, weil keine Linearität möglicher Effekte a priori unterstellt werden kann.



Die Stichprobe für die Schätzung dieses Modells muss sich natürlich auf Personen mit Haupt- oder Ko-Diagnose einer schizophrenen Störung als Beobachtungseinheit beziehen und nicht auf deren sich möglicherweise mehrfach wiederholende Behandlungsepisoden. Unter den bis zum Stichtag 1.7.1997 in allen drei Häusern registrierten 59.366 Datensätzen finden sich 12.314 Behandlungen mit Haupt- oder Nebendiagnose vom schizophrenen Formenkreis. Diese Behandlungen wurden von  $n = 5892$  verschiedenen Personen in Anspruch genommen, wobei Patientenmigration über die Häuser hinweg berücksichtigt wurde. Pro Person wurden jeweils diejenigen Informationen im nachfolgenden LISREL-Modell verarbeitet, die pro Variable frühestmöglich erhoben wurden. Das bedeutet, dass beispielsweise der erreichte Bildungsabschluss beim ersten stationären Aufenthalt verwendet wird, während die Variable ZA13 („erste psych. Auffälligkeit“), - weil sie als sogenannte „Zusatzvariable“ nicht in allen BADO-dokumentierenden Häusern erhoben wurde, - erst aufgrund eines späteren Aufenthalts an einem anderen Haus Berücksichtigung finden kann. Pro Person wird dementsprechend als Zeitpunkt der ersten Hospitalisierung (A9) das Lebensalter bei der psychiatrischen Ersthospitalisierung verwendet, unabhängig von der dabei gestellten (Erst-)Diagnose. Insgesamt konnten auf diese Weise  $n=2890$  Personen mit vollständigen Daten identifiziert werden. Die fehlenden rund 3000 Personen stammen fast ausschließlich aus Missing Values auf der Variablen ZA13 und können mit einiger Plausibilität als „missing at random“ (Little & Rubin, 1987) betrachtet werden, zumal der Zensierungsmechanismus (welche Klinik dokumentiert in welcher BADO-Form?) als unabhängig von den Personen angenommen werden kann.

Zu den empirischen Ergebnissen dieses Modells: Das Modell zeigt hervorragende globale Fit-Werte (Goodness of Fit = 1.0; Adjusted Goodness of Fit = 0.98; root mean residual = 0.060). Mit anderen Worten, die theoretische Vorstellung, dass Bildung, Geschlecht und Klinik bei schizophrenen Patienten im Zusammenhang zum „age at onset“ stehen, lässt sich mit den empirischen Daten sehr gut vereinbaren. Alle spezifizierten Pfade im Modell zeigen einen statistisch signifikanten Einfluss (alle  $|t| > 2.4$ ) auf den registrierten „onset“. Einzig der mögliche Unterschied von Klinik 2 zur Referenzklinik blieb bedeutungslos und wurde daher als „Kausalpfad“ geschlossen. Die latente Variable „onset“ kann trotz ihrer „spärlichen“ Ausstattung mit nur zwei Indikatoren mit einer sehr guten Reliabilität von  $r_{tt} = 0.89$  geschätzt werden (vgl. Abbildung unten). Das bedeutet, dass trotz der offensichtlichen Fehlerbehaftung beider Einzelvariablen ihr gemeinsamer „Kern“ bei geeigneter Spezifikation eines Messfehlermodells sehr reliabel gemessen werden kann.



Ein Vergleich mit den Ergebnissen der methodisch äußerst anspruchsvollen, bevölkerungsrepräsentativen ABC- (age beginning, course) Studie des Mannheimer Zentralinstituts f. seelische Gesundheit (z.B. Häfner et al., 1998) zum age-at-onset bei Schizophrenie zeigt, dass die Variable „ZA13“ mit einem Mittelwert von 26,73 Lebensjahren (s.d.=11,85) im Vergleich zum „Goldstandard“ dieser epidemiologischen Feldstudie etwas früher liegt (wie zu erwarten war s.u.): Der age-at-onset wurde in der ABC-Studie mittels eines standardisierten Erhebungsinstruments (IRAOS; vgl. Häfner et al., 1992) nach mehreren Kriterien bestimmt: Häfner et al. (1998) berichten das Auftreten des ersten positiven Symptoms (für Schizophrenie, nicht eine allgemeine psychopathologische Auffälligkeit) bei durchschnittlich 29,0 Lebensjahren. Die erste psychiatrische Hospitalisierung trat im Datensatz der drei Kliniken im mittleren Lebensalter von 31,63 Jahren (s.d.=12,51) auf, liegt also vergleichsweise dazu etwas

später (im Vergleich zu mittleren 30,3 Jahren bei der Ersthospitalisierung in der ABC-Studie). Dabei sind Unterschiede der regionalen Versorgungsstrukturen zwischen Mannheim und Ostbayern zu berücksichtigen.

Der Mittelwert der latenten Variable „onset“ kann aus den Gewichten der obigen Abbildung berechnet werden (durch die Schätzgleichung

$$\text{ONSET} = (1 \cdot \text{ZA13} + 0.79 \cdot \text{ERSTHOSPITAL}) / 1.79;$$

vgl. Abbildung). Er liegt in unserer Stichprobe bei 28,52 Lebensjahren und entspricht damit ziemlich genau der Schätzung der ABC-Studie für das „first positive symptom“. Wohlgedacht sind die Gewichte des hier vorgelegten LISREL-Modells nicht daraufhin optimiert worden, einen externen Mittelwert zu reproduzieren, sondern vielmehr danach, möglichst viel Varianz der latenten Variable „onset“ durch die Einflussfaktoren Bildung, Klinik und Geschlecht aufzuklären!

Weil die Varianz dieser latenten Variablen „onset“ auf den Indikator „erste psych. Auffälligkeit“ normiert wurde (Pfeil von „onset“ auf „1. psych. Auffälligkeit“ mit fixiertem Wert 1.0 in obiger Abbildung), können die beta-Koeffizienten des einer multiplen Regression entsprechenden Kausalmodells jeweils als Mittelwertsunterschiede unter simultaner Adjustierung der übrigen Einflussgrößen interpretiert werden. Das bedeutet, dass die männlichen Patienten unseres Datensatzes im Schnitt um rund 4,44 Jahre früher erkrankten als die Patientinnen, wobei mögliche Bildungsunterschiede oder Klinikdifferenzen schon berücksichtigt sind. Auch diese gefundene Effektgröße für den Geschlechtsunterschied entspricht sehr gut dem von Häfner et al. (1998) berichteten (unadjustierten) Unterschied von drei bis vier Jahren (je nach Definition des onset-Zeitpunkts).

Im Vergleich zu PatientInnen ohne formalen Bildungsabschluss sind SonderschülerInnen 4,19 Jahre früher, PatientInnen mit Hauptschulabschluss aber 8,13 Jahre später erkrankt. Bildung hat keinen linearen Zusammenhang mit dem age-at-onset: PatientInnen mit Abitur sind nur durchschnittlich 6,44 Jahre später erkrankt als solche ohne formalen Bildungsabschluss. Die spezifizierten Einflussfaktoren Bildung, Geschlecht und Klinik zusammengerechnet erreichen einen statistisch zwar hochsignifikanten, inhaltlich aber eher bescheidenen Einfluss auf den Zeitpunkt der Erkrankung: Nur 11% der Varianz der latenten Variablen „onset“ können durch die beteiligten Variablen aufgeklärt werden.

Was bedeuten diese Ergebnisse für die Validität der BADO? Der aus externen Studien vorhergesagte Geschlechtsunterschied im age-at-onset ließ sich anhand der BADO-Daten ebenfalls nachweisen, und zwar auch nach Berücksichtigung möglicher konfundierender Variablen und in einer durchaus mit anderen Studien kompatiblen Effektgröße. Die beiden BADO-gestützten Indikatoren konnten - obwohl sie jeder einzeln für sich nur mit deutlichen Messfehlern behaftet erfasst werden können - in einem geeigneten psychometrischen Modell eine sehr zufriedenstellende Reliabilität erreichen. Die hohe Übereinstimmung mit den Ergebnissen anderer Studien kann zusätzlich als inhaltliche Bewährung, d.h. als positiver Hinweis auf eine gute Validität zumindest dieses Teils der BADO-Variablen gewertet werden.

## 2. Statistische Methoden zum Krankenhausvergleich

Die Ergebnisse dieses Abschnittes wurden ausführlicher berichtet im Artikel:

Frick U., Rehm J., Krischker S., Cording C. (1999) Length of stay in a German psychiatric hospital as a function of patient and organizational characteristics – a multilevel analysis. *Int. J. of Methods in Psychiatric Research*, 8: 3, 146-161

Ein reiner Vergleich von aggregierten Kennzahlen von Krankenhäusern (z.B. ein Vergleich der durchschnittlichen Verweildauer aller PatientInnen) bedeutet für diejenigen Kliniken, die Patienten mit schwerwiegenderen Erkrankungen behandeln, einen unfairen Vergleich. Es wurden daher die in der Literatur beschriebenen statistischen Methoden zur Adjustierung des sogenannten "Patienten-Mix" aufgearbeitet und in ihrer Effizienz miteinander verglichen (vgl. auch DeLong et al., 1997). Nach den Empfehlungen von Christiansen & Morris (1997) wurde für das vorliegende Projekt der methodische Ansatz der Multi-Level-Analysen ausgewählt und schrittweise erprobt (zunächst für die Analyse der Verweildauern, aber vgl. auch Abschnitt 5).

### 2.1 Ausgangsidee

Die Verweildauer einer Behandlungsepisode wird beeinflusst von sehr unterschiedlichen Einflussgrößen, die sich zudem auf unterschiedlichen hierarchischen Ebenen ansiedeln lassen: Merkmale der aktuellen Behandlungsepisode, Patientenmerkmale, Merkmale der behandelnden Station, Merkmale der Klinik, sowie Merkmale des Behandlungssystems. Wird in einer statistischen Analyse nicht berücksichtigt, dass zum Beispiel die Merkmale einer Krankenhausstation für alle PatientInnen, die dort behandelt werden, exakt dieselben sind, und werden diese Stationsmerkmale (z.B. Bettenzahl, Personalausstattung usw.) so behandelt, als könnten sie ähnlich zu Patientenmerkmalen (Diagnose, Schweregrad, Vorerkrankungsgeschichte usw.) zwischen den Patienten frei variieren, dann wird ihr potenzieller Einfluss ggf. stark überschätzt. Denn infolge der (strukturbedingten!) Null-Variation zwischen den Patienten ein- und derselben Station wird in traditionellen statistischen Analysen für Stationsmerkmale ein Messfehler über die Personen angenommen, der aber rechnerisch Null werden muss (es gibt ja keine Variation für Stationsmerkmale innerhalb der Patienten einer Station). Da der Schätzer für den Einfluss des Stationsmerkmal an seinem geschätzten Messfehler normiert wird, wird also die Bedeutung von Stationsmerkmalen überschätzt.

In einem ersten Schritt zur Überwindung dieses weit verbreiteten Fehlers wurde im Projekt anhand der Daten des Regensburger Bezirksklinikums die Praktikabilität von hierarchischen linearen Modellen in einer Zwei-Ebenen-Analyse der Verweildauer erprobt (vgl. Frick et al., 1999). Die statistischen Grundprinzipien waren dabei folgende:

Die logarithmierte Verweildauer (length of stay = LOS) eines Patienten  $i$  wird zunächst im traditionellen Ansatz der multiplen linearen Regression als Funktion einer Vielzahl von individuellen Merkmalen (level 1 predictor variables) modelliert.

$$(1) \quad \text{LOS}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon_i$$

wobei

$i$	:=	index of individuals
$X_1 \dots X_m$	:=	level 1 predictor variables
$\varepsilon_i$	:=	individual error term

Berücksichtigt man die Tatsache, dass der zugrundeliegende Datensatz hierarchisch geordnet ist, und potenziell Stationsmerkmale (Index  $w$  für ward) von Bedeutung sein können, dann verändert sich ohne Einführung weiterer Variablen Gleichung (1) zur nachfolgenden Gleichung (2). Zusätzlich werden in (2) die individuellen Merkmale der Patienten als Abweichung vom jeweiligen Stationsdurchschnitt ausgedrückt, was die Interpretation von  $\beta_{0w}$  als mittlere Verweildauer in der Station  $w$  erlaubt.

$$(2) \quad \text{LOS}_{iw} = \beta_{0w} + \beta_1 (X_1 - \bar{X}_{1w}) + \beta_2 (X_2 - \bar{X}_{2w}) + \dots + \beta_m (X_m - \bar{X}_{mw}) + \varepsilon_{iw}$$

Die mittlere Verweildauer pro Station wird dann in einem nächsten Schritt zerlegt in eine globale Verweildauer über alle Stationen, und einen stationsspezifischen Anteil. Dies wird in Gleichung (3) ausgedrückt. Durch die Indexierung über  $w$  bedeutet das Intercept der Regressionsgleichung nun keine Konstante mehr, sondern wird selber zur Zufallsvariable, die von externen Einflüssen abhängen kann.

$$(3) \quad \beta_{0w} = \gamma_{00} + u_{0w}$$

In einer Erweiterung der Idee von stationsspezifischen Zu- und Abschlägen auf die individuellen Verweildauern kann man Gleichung (3) auch über alle  $w$  Stationen hinweg als eigene Regressionsgleichung auffassen, die die Einwirkung von Stationsmerkmalen ( $W_1$  bis  $W_k$ ) auf den jeweiligen Stationsmittelwert schätzt. Dies ist in Gleichung (3.a) erläutert.

$$(3.a) \quad \beta_{0w} = \gamma_{00} + \gamma_{01} W_1 + \gamma_{02} W_2 + \dots + \gamma_{0k} W_k + u_{0w}$$

Werden Gleichungen vom Typ (3) oder (3a) in die Gleichung (2) eingesetzt, dann kann man –geeignete Schätzverfahren vorausgesetzt– simultan den Einfluss von Patientenmerkmalen und Stationsmerkmalen auf die Verweildauer modellieren. Dabei wird der Einfluss von Stationsmerkmalen auf die Verweildauer als für alle PatientInnen der jeweiligen Station absolut gleich wirkender Zu- oder Abschlag auf die zu erwartende Verweildauer berechnet (sogenanntes "means-as-outcome model"). Als inhaltliches Beispiel kann man sich vorstellen, dass eine höhere Personaldichte (z.B. 1 zusätzlicher Arzt) auf einer Station eine bessere therapeutische Kompetenz ermöglicht, die für alle PatientInnen die notwendige Verweildauer um beispielsweise einheitlich 3 Tage verkürzen könnte.

Es ist jedoch auch vorstellbar, dass sich bestimmte Stationsmerkmale für jeden Patienten unterschiedlich auswirken können (sogenannte Kontexteffekte). Beispielsweise wirkt ein 65jähriger Patient in einer Adoleszenten-Station sehr alt, wogegen er auf einer geriatrischen Station eher jung wäre. Je nach Durchschnittsalter der MitpatientInnen (was ein level-2 Merkmal der Station darstellt), könnte sich der Einfluss des Alters auf die Verweildauer unterschiedlich auswirken. Wenn  $X_1$  in Gleichung (1)

und (2) das individuelle Alter darstellt, dann wäre eine stationsspezifische "Überformung" des Alterseinflusses  $\beta_{1w}$  ganz ähnlich wie in Gleichung (3) folgendermaßen darstellbar:

$$(4) \quad \beta_{1w} = \gamma_{10} + u_{1w}$$

Es gibt in (4) also einen über alle Personen und Stationen hinweg gleichen Einfluss des Alters auf die Verweildauer (z.B. pro Lebensjahr hypothetisch eine um einen Tag längere Verweildauer, Koeffizient  $\gamma_{10}$ ). Zusätzlich gibt es aber einen Beitrag der jeweiligen Station  $w$ , der die Steilheit des Gradienten von Alter und Verweildauer modulieren kann ( $u_{1w}$ ). Theoretisch sind dann auch Stationen vorstellbar, wo der Patient, je älter er ist, umso kürzer behandelt wird,  $u_{1w}$  also negativ und vom Betrag größer als  $\gamma_{10}$  wäre.

$$(4.a) \quad \beta_{1w} = \gamma_{10} + \gamma_{11}W_1 + \gamma_{12}W_2 + \dots + \gamma_{1k}W_k + u_{1w}$$

Analog zum Vorgehen bei Gleichung (3) kann auch für den modulierenden Effekt einer Station auf die Steigungskoeffizienten ("Random slope Model") untersucht werden, inwieweit diese stationsspezifische Überformung von individuellen Einflussgrößen auf die Verweildauer ihrerseits von Stationsmerkmalen abhängt. Dies wird untersucht in Gleichung (4.a).

Es ist zudem möglich, Effekte individueller Merkmale, Random-Slope-Effekte, und stationsspezifische Zu- und Abschläge (means-as-outcome) auf die Verweildauern simultan in einem einzigen Modell zu schätzen. Zur Durchführung der Analysen können Programme wie HLM, MLWIN, oder SAS PROC MIXED herangezogen werden.

Als Ergebnis einer statistischen Analyse aller Behandlungsepisoden aus dem Jahr 1996 in allen 20 Stationen im Akutbereich des Regensburger Bezirksklinikums (insgesamt 5840 analysierte Behandlungsepisoden) ergaben sich unter Verwendung des oben beschriebenen Multi-Level Ansatzes folgende bedeutsame Einflussgrößen auf die Behandlungsdauer.

Level-1-Predictor Variable	Random effect ?	Level-2 Predictor Variable	Coeff.	Standard Error	T value	prob (T)
Intercept ( $\beta_0$ )	yes	Intercept 2 ( $\gamma_{00}$ )	3.128	0.1405	22.26	0.000
		Mean GAF-Score ( $\gamma_{01}$ )	- 0.066	0.0275	-2.41	0.028
		Nurses' FTEs per bed day ( $\gamma_{02}$ )	- 818.538	407.7154	-2.01	0.061
		Clinicians' FTEs per bed day ( $\gamma_{03}$ )	560.294	1.508.3283	0.37	0.715
Age at admission	yes		0.005	0.0023	2.25	0.039
GAF score at admission	yes		- 0.005	0.0023	-2.28	0.037
Schizophrenic disorder (0/1) ( $\beta_3$ )	yes	Intercept 2 ( $\gamma_{30}$ )	0.052	0.0961	0.54	0.596
		Proportion of schizophrenic patients ( $\gamma_{31}$ )	1.192	0.3493	3.41	0.004
Affective disorder (0/1)	yes		0.212	0.1005	2.11	0.050
Illegal/pharmaceutical drug abuse (0/1)	yes		- 0.040	0.1642	-0.25	0.809
No complementary services after dismissal (0/1)	no		- 0.298	0.0537	-5.55	0.000
Subsequent treatment on outpatient base arranged (0/1)	no		0.106	0.0352	3.01	0.009
Improvement of GAF score	yes		0.013	0.0034	3.81	0.002
No drug treatment while hospitalized (0/1)	yes		- 0.437	0.0695	-6.28	0.000
Anti-cholinergic medication while hospitalized (0/1)	no		0.170	0.0715	2.37	0.031

## 2.2 Level-1 Prädiktoren: Einfluss von Patientenmerkmalen

Aus den zahlreichen individuellen Patientenmerkmalen, die in der BADO erfasst werden, zeigten sich nur 10 Variablen als statistisch bedeutsam, um die Verweildauer der PatientInnen vorhersagen zu können. Dabei ist zu beachten, dass die Hälfte dieser level-1-Merkmale nicht stabile "Persönlichkeitsmerkmale" (Alter, Geschlecht, etc.) oder zeitkonstante Krankheitscharakteristika (z.B. Diagnose, Erst-erkrankungsalter usw.) erfassen, sondern Merkmale des Behandlungsprozesses darstellen (z.B. Medikationsstrategie), wie er für den/die Patienten/in anberaumt wurde.

Nur drei der zehn Level-1 Prädiktoren zeigten einen über alle 20 Regensburger Stationen hinweg einheitlichen Einfluss (Random Effect = No in Table 4). Wurde ein(e) Patient(in) mit anti-cholinerger Medikation (also einem Gegenmittel gegen die akuten Nebenwirkungen von antipsychotischer Neuroleptikagabe) behandelt, dann war dies mit einer längeren stationären Behandlungsdauer verbunden. Wenn für die Zeit nach der Entlassung keine komplementärpsychiatrische Versorgung arrangiert wurde (bzw. arrangiert werden musste), dann führte dies zu einer klar verkürzten Verweildauer. Dagegen war der Übergang eines/einer Patienten/in in die hauseigene Tagesklinik zur ambulanten Weiterbehandlung meist mit einer längeren Verweildauer in der vollstationären Zeit verbunden.

Die sieben übrigen Level-1-Prädiktorvariablen entfalten ihren Einfluss auf die Verweildauer pro Station in jeweils unterschiedlicher Intensität und teilweise auch Richtung. Die in Table 4 für die Variablen

- Alter bei Aufnahme
- Maß der sozialen Anpassung bei stationärer Aufnahme
- Verbesserung der sozialen Anpassung während des Aufenthalts
- Diagnose vom schizophrenen Formenkreis
- Diagnose affektive Erkrankung
- Diagnose Medikamenten-/Drogenmissbrauch bzw. -abhängigkeit
- Behandlung ohne Psychopharmaka

dargestellten Koeffizienten bilden den über alle 20 Stationen des Regensburger Klinikums gemittelten durchschnittlichen Einfluss, der der jeweiligen Variablen zuzurechnen ist. So verlängert im Schnitt jedes Lebensjahr der PatientInnen den stationären Aufenthalt um 0.005 Log-Tage (die abhängige Variable wurde ja vor der Analyse log-transformiert). Dieser Alterszuschlag wirkt sich in manchen Stationen deutlich steiler aus, in anderen Stationen dagegen erheblich flacher. Ebenso bedeutet die Diagnose einer schizophrenen Erkrankung eine im Durchschnitt um 0.052 log-Tage längere Verweildauer, bzw. bei einer affektiven Erkrankung eine um 0.212 log-Tage verlängerte Verweildauer (im Vergleich zu den übrigen PatientInnen). Die genaue statistische Erwartung dieses Diagnosen-Verlängerungseffekts hängt jedoch auch von der Station ab, die über die Entlassung zu befinden hat. PatientInnen mit Suchterkrankungen (Medikamente oder illegale Drogen) sind im Durchschnitt in kürzerer stationärer Behandlung als die übrigen PatientInnen (Entgiftung!), wobei auch hier die Station diesen Verkürzungseffekt mit determiniert.

Ebenso bedeutet die Tatsache einer Behandlung ohne den Einsatz von Psychopharmaka, dass im Prinzip für diese PatientInnen mit einer deutlich reduzierten Verweildauer (-0.437 log-Tage) zu rechnen ist. Wie stark dieser Effekt jedoch ausfällt, hängt von der Station ab. Stationäre Aufenthalte ohne den Einsatz von Psychopharmaka finden vor allem statt bei Alkoholkranken, die oft im Rausch stationär aufgenommen werden, und die Klinik ohne weitere Entgiftung sehr bald wieder verlassen. Oder sie finden statt auf Psychotherapie-Stationen, wo neurotische Erkrankungen oder Persönlichkeitsstörungen ohne den Einsatz von Psychopharmaka therapiert werden. Für letztere PatientInnen ist die Behandlungsdauer deutlich länger als für die Alkoholkranken mit 1-Tages-Verweildauern. Daher macht es Sinn, den Verweildauer-Verkürzungseffekt der Psychopharmaka-freien Behandlung pro Station gesondert zu berechnen.

Das Ausmaß der sozialen Anpassung (GAF-Score) bei Aufnahme verkürzt im Schnitt die Therapiedauer um 0.005 Log-Tage pro GAF-Punkt. Je nach Station wird der Einfluss der Aufnahme-GAF aber unterschiedlich bei der Bestimmung der notwendigen Behandlungsdauer berücksichtigt. Die durch den stationären Aufenthalt erzielte Verbesserung in der sozialen Anpassung wurde gemessen als Differenz der Entlassungs-GAF und der Aufnahme-GAF. Je mehr "Profit" (im Sinne einer Verbesserung der GAF-Werte) ein Patient / eine Patientin durch den stationären Aufenthalt erzielt, umso länger hat dieser Aufenthalt im Durchschnitt dann auch gedauert. Es spielt dabei keine Rolle, von welchem Ausgangsniveau aus der/die Patient(in) diesen Zuwachs an sozialer Anpassung erlebt, denn für die unterschiedlichen GAF-Werte bei Aufnahme wurde im statistischen Modell bereits adjustiert. Sehr wohl eine Rolle spielen allerdings die Stationen. M.a.W., der Zugewinn von sozialer Anpassung ist auf manchen Stationen "mühsamer", -d.h. mit längeren Verweildauern verbunden, - zu erreichen als auf anderen. Dies ist umso bedeutsamer für den Vergleich der Stationen, als für die individuellen Merkmale der PatientInnen (Alter, Diagnose, Erkrankungsdauer, Sozialstatus, usw.) ja schon adjustiert wurde.

### **2.3 Level-2 Prädiktoren: Einfluss von Organisationsmerkmalen**

Neben den oben bereits erwähnten Überformungen von individuellen Patientenmerkmalen (random slopes), die deren Einfluss auf die Verweildauer je nach Maßgabe der über die Entlassung entscheidenden Station beeinflussen, konnten für das Bezirksklinikum Regensburg noch weitere Stationsunterschiede und -merkmale mit Bedeutung für die Verweildauer gesichert werden.

Das Random-Intercept im endgültig akzeptierten statistischen Modell bedeutet, dass auch nach Berücksichtigung aller Patientenmerkmale durch die Stationen unterschiedlich lange Verweildauern gepflogen werden, die nicht über Patientenunterschiede erklärbar sind. Frick, Rehm et al. (1999) haben den Anteil, den diese "Stationsgewohnheiten" innerhalb von Regensburg an der Gesamtvarianz der Verweildauern verursachen, mit knapp 28% berechnet. Es ist dieser Anteil an organisationsbedingten Verweildauer-Unterschieden, an den sich qualitätssichernde oder qualitätsverbessernde Maßnahmen richten können, wenn eine Vereinheitlichung der Verweildauer unter Wahrung der von Patientenseite her notwendigen Behandlungsbedürfnisse stattfinden sollte. Notabene: 28% Potenzial zur Vereinheitlichung der Verweildauer heißt nicht, dass um 28% kürzere Verweildauern erreichbar wären!

Der entscheidende Vorteil der hier gewählten Methode zum Vergleich von Institutionen liegt darin, dass nicht nur ein "fairer" Vergleich der Verweildauern über die Stationen hinweg möglich ist. Über diesen "fairen" (sprich: für Patientenunterschiede adjustierten) Vergleich hinausgehend können auch noch Charakteristika der Organisationen selber untersucht werden, ob und wie stark sie Einfluss auf die nach Adjustierung von Patientenunterschieden noch bestehenden Stationsunterschiede in der Verweildauer ausüben. Darin liegt das Neuerungspotenzial von Multi-Level-Analysen beim Vergleich von Krankenhäusern: Während traditionelle Prozeduren zum Risk-Adjustment (vgl. Iezzoni et al., 1996) lediglich darauf fokussieren, Unterschiede zwischen den Krankenhäusern unter Wahrung eines fairen Vergleichsmaßstabs statistisch abzusichern, vermag eine hierarchische Analyse (Multi-Level-Analyse) zusätzlich, die organisatorischen Merkmale der Krankenhäuser zu bestimmen, welche die Unterschiede erklären können. Damit wird dann im Anlassfalle auch klar, an welchen "Stellrädern" mögliche organisatorische Veränderungen anzusetzen haben, wenn eine Verbesserung der Behandlungsergebnisse, der Effizienz, oder der Kosten erstrebt werden soll.

Am Beispiel des Stationsvergleichs im Regensburger Klinikum wurden mehrere Variablen untersucht, die für solche organisationsbedingten Unterschiede in den Behandlungsgewohnheiten (Verweildauern) verantwortlich sein können:

- Die **Größe der jeweiligen Station** (gemessen über die Bettenzahl, Zahl der erbrachten Pflegetage im Beobachtungszeitraum, Anzahl der behandelten PatientInnen, sowie den Nutzungsgrad im Beobachtungszeitraum),
- ihr spezifisches **Behandlungsprofil** (gemessen über den Anteil, den jeweils schizophrene, depressive, alkoholranke oder drogenabhängige PatientInnen am gesamten Patientenaufkommen der Station ausmachen),
- das **soziodemographische Profil** der Stationen (Anteil der Geschlechter, Durchschnittsalter der PatientInnen, durchschnittlicher Eingangs-GAF-Score der PatientInnen),
- sowie die **Personalausstattung** (Verhältnis von pflegerischen bzw. therapeutischen Vollzeitstellenäquivalenten zu insgesamt erbrachten Pflegetagen im Beobachtungszeitraum)

wurden daraufhin untersucht, ob sie die für das Random-Intercept beobachtbaren Stationsunterschiede statistisch bedeutsam aufklären können (vgl. Gleichung (3.a) im Abschnitt 2.1). Drei der genannten Variablen konnten dabei als signifikant beeinflussend gesichert werden:

Je deutlicher in einer Station PatientInnen mit einem hohen sozialen Funktionslevel das Stationsbild dominierten (hoher mittlerer Eingangs-GAF der PatientInnen), umso kürzer verliefen auch die anschließenden Behandlungen. Dieser Effekt ( Koeffizient  $\gamma_{01}$  in Table 4 = -0.066 ) ist kein Effekt auf der Personen-Ebene, sondern ein klarer Organisations-Effekt. Für die individuell pro Patient aufgrund seiner guten sozialen Funktionsfähigkeit verringerte Behandlungserfordernis wurde im oben dargestellten statistischen Modell ja bereits adjustiert. Es sind zwei (einander keineswegs ausschließende) Deutungen dieses Effekts möglich: Entweder entwickelt sich in Stationen mit hohem Durchschnitts-GAF ein besonders günstiges therapeutisches Klima, das sich in der Folge real-inhaltlich Verweildauer-verkürzend auswirken kann. Umgekehrt formuliert bedeutet dies dann: In Stationen mit geringem Eingangs-GAF der PatientInnen wirkt sich der ungünstige PatientInnen-Mix auf die dort aufgenommenen PatientInnen zusätzlich Verweildauer-verlängernd als Nachteil aus. Oder, -und dies stellt eine alternative, eher sozialpsychologisch orientierte Erklärung dar, - die Ärz-

Innen, die in solchen Stationen mit durchschnittlich sehr hohem sozialen Funktionsniveau über die Entlassungsfähigkeit eines/einer Patienten/in zu entscheiden haben, entwickeln vor dem Hintergrund der guten sozialen Anpassung ihrer PatientInnen einen "optimistischen Bias", der auch PatientInnen mit eher ungünstigerem Therapieverlauf zu einem einheitlichen Verkürzungs-Bonus bei der Entlassung verhilft.

Bezüglich der Personalausstattung in den 20 untersuchten Stationen im Klinikum Regensburg konnte kein Zusammenhang der Personaldichte des therapeutischen Personals (gezählt: Ärzte, Psychologen, Sozialarbeiter) mit der Verweildauer gesichert werden. Für das Pflegepersonal hingegen zeigte sich ein möglicher Zusammenhang in folgender Richtung: Je mehr Pflegepersonal pro stationsweise erbrachtem Behandlungstag zur Verfügung stand, umso kürzer verliefen auf solchen Stationen im Schnitt die Behandlungen. M.a.W., je geringer die Zeitknappheit des Pflegepersonals gegenüber den anwesenden PatientInnen war, umso schneller verlief die Behandlung. Die Tatsache, dass ein ähnlicher Effekt des intensivierten Personaleinsatzes für das therapeutische Personal (Ärzte, Psychologen, Sozialarbeiter) nicht sicherbar war, ist im konkreten Datensatz vor allem darauf zurückzuführen, dass die Verteilung der Stellenäquivalente über die Stationen und Behandlungszahlen bei den Ärzten und Psychologen deutlich gleichmäßiger organisiert war als beim Pflegepersonal. Wo aber in der unabhängigen Variable (physicians' FTEs per bed day) keine Varianz besteht, kann diese auch keine Kovarianz mit der abhängigen Variable (Verweildauer der Patienten) erreichen.

Das diagnostische Spektrum einer Station als indirektes Merkmal des dort ablaufenden therapeutischen Geschehens wurde gemessen über die prozentualen Anteile, die die vier Diagnosengruppen (Schizophrenie, Depression, Alkohol, Drogen) am Gesamtklientel einer Station ausmachen. In Table 4 wird der Zusammenhang, der für die individuelle Diagnose "Schizophrenie" mit der Verweildauer besteht, als vom jeweiligen Stationsumfeld mitbestimmt ausgewiesen (random effect = yes). Dabei konnte nachgewiesen werden, dass sich vor allem diejenigen Stationen durch besonders lange Schizophrenie-Behandlungen auszeichnen, die auch besonders spezialisiert nur PatientInnen dieses Formenkreises therapieren. Je mehr schizophrene MitpatientInnen auf einer Station anzutreffen sind, umso stärker (Koeffizient  $\gamma_{31} = 1.192$ ) wirkt sich die individuelle Diagnose "Schizophrenie" für den Patient / die Patientin Verweildauer-verlängernd aus. Dabei muss allerdings beachtet werden, dass "auswirken" im Sinne einer statistischen Assoziation, und nicht im Sinne einer kausalen Verursachung nachgewiesen wurde. Spezialisierte Schizophrenie-Stationen bedeuten also für schizophrene Patienten eine höhere Verweildauer, als sie allein aufgrund der Diagnose zu erwarten gewesen wäre. Wiederum muss bedacht werden, dass für sämtliche weiteren in der BADO gemessenen Patientenvariablen, so sie bedeutsam mit der Verweildauer korrelierten, schon rechnerisch adjustiert wurde.

Daher trifft zur Erklärung dieses Kontext-Effektes der Schizophrenie-Verweildauerverlängerung besonders in Spezialstationen eine mögliche Patienten-Selektion ("hier wurden prognostisch besonders ungünstige PatientInnen zugewiesen") nicht zu. Analog zum Kontext-Effekt bei den mittleren GAF-Werten pro Station kann auch bei Schizophrenie-Spezialstationen ein ko-therapeutischer Effekt durch PatientInnen aus anderen Erkrankungsformen (zumeist wird mit depressiven Erkrankungen gemischt) als Erklärung auf der realen Wirkungsebene gedacht werden. Werden PatientInnen mit schizophrenen Erkrankungen nur unter ihresgleichen behandelt, werden sie möglicherweise von diesem ko-therapeutischen Effekt depriviert. Aber auch die sozi-

alpsychologische Erklärung ausschließlich auf der Urteilebene kann nicht ausgeschlossen werden: Ärzte, die auf spezialisierten Schizophrenie-Stationen über die Entlassung ihrer PatientInnen zu entscheiden haben, wären danach wegen der Häufung der prognostisch eher schwierigeren Situation ihrer PatientInnen mit der Entlassung noch vorsichtiger, als sie dies allein aufgrund der Diagnose sein müssten (unter Einbeziehung aller übrigen Patientenmerkmale). Es handelt sich dann bei den verlängerten Aufhalten um einen Urteilsfehler vom Muster eines Assimilierungsprozesses (Schwarz & Bless, 1992).

Alle übrigen Stationsmerkmale, also Größe der Station, diagnostische Spezialisierung bei anderen Diagnosen, Geschlechts- und Altersgliederung auf den Stationen konnten sich nicht als Prädiktoren der Verweildauer auf der Ebene der Organisationsmerkmale durchsetzen. Während das Alter aber als individuelles Merkmal (Level 1) sehr wohl von Bedeutung war, spielte das Geschlecht der PatientInnen weder als individuelles Merkmal, noch als Merkmal der Stationsgliederung eine Rolle bei der Behandlungsdauer: Frauen wurden weder in gemischten Stationen, noch in reinen Frauenstationen länger oder kürzer behandelt als Männer mit vergleichbarer Erkrankung, vergleichbarem Alter, Schweregrad etc. .

Zusammenfassend zeigte sich aus den im Abschnitt 2 dargelegten Analysen, dass mit der Wahl von Multi-Level-Analysen als Werkzeug zum Krankenhausvergleich ein entscheidender Durchbruch gelingen konnte: Einerseits werden die Behandlungsergebnisse (hier: Verweildauer) als faire Vergleiche ohne Verzerrung durch unterschiedlich schwer erkrankte PatientInnen einander gegenüber gestellt. Es werden also nicht fälschlich Krankenhäuser oder Stationen identifiziert, deren "bessere" oder "schlechtere" Ergebnisse (egal auf welchem Bewertungsmaßstab) sich schon allein durch ihren spezifischen Patienten-Mix erklären lassen. Auch für dichotome Zielkriterien (z.B. Wiederaufnahme binnen 3 Monaten nach Entlassung), oder für zensierte Daten (z.B. Zeit bis zur Wiederaufnahme), die möglicherweise besser als die reine Verweildauer geeignet sind, Qualitätsaspekte von Krankenhäusern zu erfassen, wurden entsprechende Adaptationen des statistischen Modells von hierarchischen linearen Modellen bereits entwickelt und publiziert.

Andererseits können für Institutionen, die in ihren Ergebnissen bei Multi-Level-Analysen als von der statistischen Erwartung abweichend auffallen, durch den Einbezug von level-2 Merkmalen dann auch Hinweise gegeben werden, in welche Richtung ggf. organisatorische Änderungen erfolgen sollten.

Leider ist in der politischen Realität der deutschen Versorgungslandschaft dieses Erkenntnis, dass nur Analysen, die den hierarchischen Aufbau der Entlassungsstatistiken der Krankenhäuser (Patienten in Stationen in Krankenhäusern mit Merkmalen auf jeder der drei Datenebenen) berücksichtigen, zu stimmigen Ergebnissen führen können, noch nicht berücksichtigt (vgl. Cording et al., 2001). Die bisherigen Vereinbarungen zur Umsetzung von Krankenhausvergleichen nach §5 Bundespflegegesetzverordnung bewegen sich ausschließlich auf der Ebene aggregierter Daten (unter Ausschluss individueller Merkmale und Datensätze) und unterliegen damit einem erhöhten Risiko von "ökologischen Fehlschlüssen" (vgl. von Korff et al., 1992)<sup>1</sup>.

---

<sup>1</sup> Ein Beispiel für einen ökologischen Fehlschluss: Aus der Tatsache, dass in Stadtteilen mit erhöhter Arbeitslosigkeit bei Wahlen die SPD immer besonders gut abschnidet, wird geschlossen, dass Arbeitslose verstärkt SPD wählen. Dies kann, muss aber nicht der Fall sein. Ebenso möglich wäre eine Erklärung, dass in Stadtteilen, wo

### 3. Stichprobe eines beispielhaften Krankenhausvergleiches

Im Rahmen dieses Forschungsvorhabens erklärten sich die psychiatrischen Kliniken in

- Augsburg,
- Jena,
- Landshut,
- Mainkofen,
- Münster,
- Osnabrück,
- Regensburg und
- Wiesloch

dazu bereit, die anonymisierten Datensätze der jeweils gemäß den Empfehlungen der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN, vgl. Cording et al., 1995) geführten Basisdokumentationen (BADO) aus den Akutversorgungsbereichen aller Häuser zusammenzuführen. Einschlusskriterium für die Aufnahme eines Datensatzes in die Vergleichsstudie war eine Entlassung im Bezugsjahr 1998. Bei zwei der acht Kliniken musste aus Gründen der Datenverfügbarkeit das Bezugsjahr 1997 gewählt werden. Die Stichprobengröße im deskriptiven Teil dieses Berichts beträgt insgesamt  $N = 28'112$  Behandlungen. Ausgeschlossen wurden im Vorfeld diejenigen Datensätze aus insgesamt sechs Stationen (in zwei betroffenen Kliniken), bei denen über das gesamte Bezugsjahr hinweg jeweils weniger als 10 Entlassungen registriert worden waren ( $n=52$ ). Für diese Stationen ergibt sich eine zu geringe Fallzahl, als dass für sie die Hierarchie des Datensatzes zuverlässig hätte berücksichtigt werden können.

Für weitere 320 PatientInnen fehlten trotz detaillierter Versuche zur Imputation von fehlenden Werten die im Abschnitt 4 dieses Berichts diskutierten Angaben zu Kernvariablen der PatientInnen. Dadurch verringerte sich bei den Analysen im nachfolgenden Abschnitt 4 die Stichprobengröße um 1,13% auf 27.792 Patienten, die in 135 Stationen (ohne die oben erwähnten sechs) behandelt worden waren. Eine Vollständigkeitskontrolle der jeweiligen Basisdokumentationen der acht Häuser konnte im Rahmen dieser Studie zwar nicht vorgenommen werden; aber da die DGPPN-BADO jeweils auf Abrechnungsdaten der Klinikverwaltungen aufbaut, gehen wir von einer weitgehend vollständigen Erfassung aller PatientInnen aus. Die Organisationsgröße der beteiligten Kliniken schwankte zwischen minimal 6 Stationen und maximal 25 Stationen. Ihre Behandlungskapazität betrug zwischen 1266 und 6583 PatientInnen, welche im Bezugsjahr aus der jeweiligen Klinik entlassen worden waren.

#### 3.1 Datenabgleich

Ein Gutteil des Arbeitsaufwandes im vorliegenden Forschungsprojekt musste dazu eingesetzt werden, die aus den acht erwähnten Häusern zugesagten Vergleichsdaten über unterschiedliche Software-Systeme hinweg und aus leicht voneinander ab-

---

Armut besonders augenfällig wird, sich die "bürgerliche Mitte" aus individuellen Motiven (z.B. Gerechtigkeitsideal) heraus verstärkt für die SPD entscheidet.

weichenden Modifikationen der eigentlich standardisierten DGPPN-BADO heraus einander kompatibel zu gestalten.

Die Teilnehmer haben teils aus eigener Initiative, teils auf Anregung durch den Projektleiter freiwillig an der Untersuchung teilgenommen. Mit zwei Partnern arbeitet die AG Versorgungsforschung schon seit 1995 bei informellen Klinikvergleichen mittels der BADO zusammen. Die weiteren Teilnehmer haben die BADO teils schon seit 1995, teils aber auch erst seit 1997 eingeführt. Es kann als gesichert gelten, dass die teilnehmenden Kliniken eine Positivselektion darstellen hinsichtlich der Motivation zur umfassenden Dokumentation der eigenen Patientendaten und hinsichtlich der Bereitschaft zur Transparenz im kollegialen Austausch.

Vier der untersuchten Kliniken stammen aus Bayern, je eine aus Thüringen, Baden-Württemberg, Niedersachsen und Nordrhein-Westfalen. Mit Ausnahme einer Klinik verfügen alle Kliniken über ein definiertes Pflichtversorgungsgebiet, davon eine als Universitätsklinik. Die universitäre Psychiatrie ist also kaum, die Abteilungspsychiatrie gar nicht vertreten, was daran liegt, dass in psychiatrischen Abteilungen die BADO meist nicht eingeführt ist und damit keine vergleichbaren Daten vorliegen. Das Übergewicht der bayerischen Kliniken ist einerseits aus der traditionell guten Zusammenarbeit zwischen diesen Kliniken zu erklären, andererseits aber auch die Folge einer sehr frühen (seit 1990 flächendeckenden) Einführung der BADO in den bayerischen Bezirkskrankenhäusern.

Obwohl die teilnehmenden Kliniken sehr interessiert an einem Krankenhausvergleich waren, und wir mit unserem Forschungsprojekt auf eine große Hilfsbereitschaft bei der Zusammenführung der Datenbanken stießen, erwies sich die Datensammlung wider Erwarten als sehr aufwändig.

Zum einen gibt es in einigen Häusern erhebliche Verzögerungen von bis zu einem Jahr bei der Dateneingabe. Dies war in einigen Kliniken Folge einer Umstellung der klinikweiten EDV-Anlage auf ein Klinikinformationssystem (KIS), das auch die BADO in einer zentralen Datenbank integrieren sollte. Dieses Vorhaben scheiterte aber. Zum anderen war der bedeutsame zeitliche Verzug in der BADO-Datenerfassung in einigen Häusern begründet durch den Einsatz von Halbtageskräften und/oder Sekretärinnen, die diese Aufgabe als „Zwischendurch-Beschäftigung“ mit erledigen sollten. Kam es bei dieser Arbeitsorganisation zu urlaubs- und/oder krankheitsbedingten Ausfällen, kulminierten die nicht erfassten BADO-Bögen. Daher ist es uns nicht gelungen, von allen acht Kliniken vollständige Daten aus demselben, einheitlichen Kalenderjahr zu benutzen. Für sechs von 8 Kliniken bildete daher das Bezugsjahr zum Klinikvergleich das Jahr 1998, für 2 Kliniken sind die Daten aus 1997 das Referenzjahr.

### **3.1.1 Datengewinnung**

Zunächst wurden die Klinikdirektoren der an einem Vergleich interessierten Kliniken angeschrieben. Sie erhielten einen Vertrag, mit dem Sie das BKR mit der Analyse der Daten beauftragten und der den beteiligten Häusern bei der Veröffentlichung von Ergebnissen Anonymität zusicherte.

Nach Vertragsunterzeichnung setzte sich die Arbeitsgruppe mit dem zuständigen BADO-Beauftragten in Verbindung, um die technischen und organisatorischen Gegebenheiten vor Ort zu prüfen. Alle Kliniken gaben zwar an, mit der DGPPN-BADO zu arbeiten, doch konnte im Detail weder technisch noch semantisch von einheitlichen Standards gesprochen werden:

Drei Kliniken arbeiteten schon mit KIS-Systemen mit dezentraler Dateneingabe und verwendeten die DGPPN-BADO (1995). Zwei Kliniken stellten gerade Ihre „alte“ BADO (DOS-Version) auf die „neue“ Dokumentation (Windows-Version) um. Zwei Kliniken arbeiteten mit der alten BADO, eine Klinik hatte bereits auf die neue BADO umgestellt.

Alle Daten wurden aus Sicherheitsgründen persönlich vor Ort abgeholt, wobei die Daten entweder über Netzwerk oder eine CD auf einen Laptop überspielt und vor Ort im Beisein des jeweiligen BADO-Beauftragten anonymisiert wurden. Alle Daten wurden zusätzlich auf Viren gescannt. Den höchsten Aufwand bei der Aufbereitung der Daten mussten die Kliniken mit KIS investieren, weil dort die BADO-Daten nicht in tabellarischer Form vorlagen und in mehrtägiger Arbeit aus getrennten KIS-Tabellen (mit Funktionalität für das System des jeweiligen Hauses) in eine einheitliche BADO-Tabelle überführt werden mussten. Bei den Kliniken, die gerade auf das neue Datenformat der DGPPN-BADO umgestellt hatten, war es beim Übersetzen der Daten von Alt auf Neu zu erheblichen Pannen gekommen, die erst durch die Plausibilitäts-Checks im Rahmen unseres Forschungsprojektes aufgedeckt wurden. Das führte in der Folge zu erheblichen Verzögerungen bei der Datenzusammenführung, da die Software-Firma das Transformationsprogramm für ihre beiden Programmversionen erst neu schreiben musste. Erst nach der erfolgreichen Implementation dieser BADO-Transformation war es uns in den betroffenen Häusern schließlich doch noch möglich, BADO-Daten aus dem Bezugsjahr abzuholen und erneut für die gepoolte Analyse aufzubereiten.

Neben der Unterstützung für das Pooling bereits bestehender, Patienten bezogener EDV-Datensätze als Grundlage für freiwillige Klinikvergleiche wurden die BADO-Beauftragten aller Häuser darum gebeten, alle verfügbaren schriftlichen Unterlagen (insbesondere Glossare, Hinweise zur Verwendung der BADO-Bögen) über die BADO zur Verfügung zu stellen und eine Checkliste auszufüllen, in der nach Codierungsgewohnheiten und Besonderheiten im Data-Flow gefragt wurde (Suizid, Fixierungen, Tötlichkeiten, geschätzte Zeit bis Bogenrücklauf etc.).

### **3.1.2 Datenaufbereitung und -prüfung**

Nach dem Einlesen der Daten in die Regensburger SAS-Datenbank wurden mit einem eigens geschriebenen Syntax-File einheitliche Variablennamen vergeben und neue Variablen erzeugt, z.B. die Klinik-ID, kumulierte Verweildauern oder die Differenzen zwischen Aufnahme- und Entlassungs-GAF-Wert. Datenfelder mit Mehrfachantworten wurden in einzelne Datenfelder zerlegt, die jeweils mit 0 und 1 codiert wurden.

Die Daten wurden auf syntaktische und inhaltliche Übereinstimmung mit der DGPPN-BADO geprüft. Das geschah zunächst mit einer Häufigkeitszählung über alle vorhandenen Variablenausprägungen. Dabei wurde schnell klar, welche Variablen gar nicht

vorhanden waren, welche Antwortkategorien nur agglutiniert (im Vergleich zum Standard der DGPPN) oder aber feiner differenziert erfasst worden waren und welche numerischen Ausprägungen zwar vorhanden waren, semantisch aber eine andere Bedeutung hatten. Bei einer Klinik wurden z.B. Suizidversuche als zusätzliche Antwortkategorie unter der Variable „Suizid“ codiert, was zu einer abnorm hohen Zahl von „Suiziden“ geführt hätte.

Die semantische Konsistenz der Daten untereinander wurde durch ein eigens geschriebenes „Plaus-Check-Programm“ getestet, das offensichtliche Codierungsfehler schnell aufdecken konnte. Beispiele für solche Plausibilitätskontrollen sind z.B. logische Abhängigkeiten zwischen Variablen (Einkommen = „Altersrente“ und gleichzeitig Aufnahmealter < 60 Jahre). Die so entstehenden Fehlermeldungen wurden den jeweiligen Kliniken mit der Bitte um Prüfung bzw. Korrektur rückgemeldet und die Daten entsprechend korrigiert.

Die Daten durchliefen drei Stufen von Plausibilitätschecks:

- Häufigkeitsauszählung,
- logische Konsistenz und
- Prüfung durch die jeweilige Klinik.

Parallel dazu wurden die Glossare bzw. schriftlichen Unterlagen aus den Kliniken sowie die Checklisten zu Besonderheiten im Umgang mit den BADO-Daten auf Übereinstimmung mit dem offiziellen DGPPN-Glossar geprüft.

Diesem ersten, pro Klinik individualisierten Schritt der Datenaufbereitung folgte die Zusammenfassung aller Klinikdaten in einer gepoolten Datentabelle mit erneuter Häufigkeitsauszählung, diesmal im Vergleich über alle Häuser. Falls in der Auszählung statistisch völlig unplausible Abweichungen bei einzelnen Kliniken auffielen und der Verdacht auf voneinander abweichende Kodiergewohnheiten bestand, wurden die Kliniken erneut um eine Stellungnahme bzw. Korrektur gebeten.

In denjenigen Fällen, wo die Antwortkategorien zwischen den Kliniken voneinander abwichen, wurde in der Regel ein gemeinsames Antwortschema für alle Häuser in der Weise entwickelt, dass die Daten auf den „kleinsten gemeinsamen Nenner“ konvertiert, d.h. differenzierte Ausprägungen zu einer gemeinsamen Kategorie zusammengefasst wurden. Der Verzicht auf Information war notwendig zugunsten eines einheitlichen Vergleichsmaßstabes.

### **3.1.3 Verbesserte Datenerhebung in den Kliniken**

Um die Möglichkeiten einer verbesserten Datenerhebung zu überprüfen, haben wir in einem halbstandardisierten Interview innerhalb des Klinikums Regensburg die Assistenzärzte (je 2 Anfänger (0-1 Jahr), Kollegen mit mittlerer Erfahrung (1-3 Jahre) und „alte Hasen“ (>3 Jahre Erfahrung mit der BADO)) bei der Bearbeitung der Erfassungsbögen per Video beobachtet und sie zu „lautem Denken“ (think aloud method) angeregt. Dabei zeigte sich, dass die BADO-Bögen unübersichtlich und die Bedeutungen mancher Variablen nicht bekannt waren. Die Arbeitsgruppe Versorgungsforschung am BKR hat daraufhin die BADO-Bögen überarbeitet und vom Layout her übersichtlicher gestaltet. Die Variablengruppen werden jetzt besser wahrgenommen, Sprungmarken sind klarer. Ein verbessertes Glossar, das jedem Arzt persönlich im

„Kitteltaschenformat“ überreicht wird, hilft zum Verständnis der Bedeutung der Variablen. Neue Kollegen sollten nach unseren Projekterfahrungen von den jeweiligen BADO-Abteilungen aktiv angeschrieben werden, mit den wichtigsten Informationen versorgt und zu einem Info-Gespräch eingeladen werden. Explizite Schulungsmaßnahmen sind für den Einsatz im Routinebetrieb, besonders wenn es um die Anwendung im gesamten Bundesgebiet geht, zwar sicherlich wünschenswert, aber im Anlassfalle wohl kaum flächendeckend zu realisieren.

Allen Herstellern von BADO-Software bzw. KIS-Anbietern haben wir den DGPPN-Merkmalkatalog mit allen Kodierungsanleitungen sowie die im vorliegenden Projekt entwickelten Plausibilitätsregeln zur Verfügung gestellt. Damit soll eine einheitliche Kodierung der BADO-Daten gewährleistet und die semantische Korrektheit verbessert werden.

### **3.1.4 Strukturanalyse**

Dass Verweildauer und andere Ergebnisvariablen nicht nur durch Patientenmerkmale, sondern auch durch andere Einflüsse wie Personalausstattung oder komplementäre Versorgung beeinflusst werden, wurde in Auswertungen bislang kaum systematisch berücksichtigt (vgl. aber Frick & Laschat, 1996). Die Auswertungen bewegen sich meist auf Patientenebene. Wir wollten dagegen auch die Strukturebenen der Stationen und der Krankenhäuser sowie deren Umfeld in die Auswertung mit einbeziehen

Deshalb haben wir – aufbauend auf den Vorgaben des Zentralinstitutes für Seelische Gesundheit in Mannheim und unter Berücksichtigung des KTQ-Fragebogens - einen Strukturfragebogen entwickelt und erprobt, der in verschiedenen Modulen (Station, Abteilung, Pflegedienst, Sozialdienst, Verwaltung) die Charakteristika einer Klinik und des Umfelds erfasst.

- Der Bogen für den Oberarzt (Abteilung) erhebt zum Beispiel Dienstjahre und Ausbildung (Facharzt, Zusatzausbildungen) aber auch die Besonderheiten der vom jeweiligen Oberarzt geführten Stationen.
- Der Stationsfragebogen wiederum erfasst z.B. die Personalstärke und die Qualifikation der Assistenzärzte.
- Besondere Bedeutung erhält der Bogen für den Sozialdienst, da der Sozialdienst die komplementäre Versorgung im Einzugsbereich der jeweiligen Klinik wohl am besten kennt.

Leider sahen sich nur vier Kliniken in der Lage, den Fragebogen auszufüllen. Die Angaben aus drei Kliniken waren dabei allerdings so lückenhaft, dass sich daraus keine auswertbaren Datensätze ergaben. Es bleibt zu hoffen, dass im Rahmen des KTQ-Verfahrens in Zukunft regelmäßig aktualisierte Strukturdaten aus den Krankenhäusern zur Verfügung stehen.

### **3.1.5 Datenschutz**

Der Umgang mit Patientendaten und damit auch Daten für Forschungszwecke wird durch eine Reihe von Gesetzen geregelt. Das StGB schützt die Patientendaten umfassend (§203 StGB), die Landeskrankengesetze modifizieren diesen umfas-

senden Schutz dahingehend, dass Patienten im Krankenhaus ohne Kommunikationsprobleme behandelt werden können. Auch die Fortbildungs- und Forschungszwecke der Krankenhäuser werden berücksichtigt (vgl. Art.27, Abs.4 BayKrG). Landes- und Bundesdatenschutzgesetze bilden dazu den Rahmen (vgl. §3 BDSG). Daneben gibt es noch eine Reihe spezieller Regelungen und Gesetze (z.B. Krebsregistergesetz), die Forschungsvorhaben auf speziellen Gebieten (z.B. Epidemiologie) erleichtern sollen.

Dokumentationspflicht und die gesetzliche Vereinbarungen zur Qualitätssicherung im Gesundheitswesen (z.B. §137 SGB V) verpflichten aber auch zur Erfassung und Verarbeitung von Daten, die nicht unbedingt der direkten Behandlung des Patienten dienen, wohl aber zur Verwaltung eines Krankenhauses notwendig sind. Obwohl durch §137 SGB V und den Auftrag des BMG an das BKR zur Durchführung des Projekts „Freiwilliger Krankenhausvergleich“ das öffentliche Interesse an Krankenhausvergleichen deutlich bekundet wird, ist klar, dass dieses öffentliche Interesse nicht die Durchbrechung der Schweigepflicht rechtfertigt, wenn personenbezogene Daten verarbeitet werden sollen. Die Bedingungen des Datenschutzes müssen eingehalten werden, ohne wichtige Projektziele zu gefährden.

Für die Verarbeitung personenbezogener Daten ohne die explizite Einwilligung der Betroffenen bildet innerhalb Bayerns Art.27 BayKrG eine angemessene Grundlage, weil die Daten innerhalb eines Krankenhauses auch ohne Einwilligung von Patienten zur Forschung genutzt werden können. Daneben gestattet Art.27, Abs.4 auch die externe Verarbeitung der Daten, sofern ein entsprechender Auftrag vorliegt und die verarbeitende Stelle ebenfalls ein Krankenhaus ist. Nach Art.27, Abs.5 ist eine Übermittlung von Patientendaten auch zulässig, wenn eine Rechtsvorschrift das erlaubt oder wegen „verwaltungsmäßiger Abwicklung“. Zur verwaltungsmäßigen Abwicklung zählt wohl auch der Auftrag zur Qualitätssicherung nach §137 SGB V.

Dennoch gilt: Anonymisierung der Daten ist der beste Schutz vor Missbrauch. Für die Zusammenführung der Daten aus verschiedenen Krankenhäusern, insbesondere bundesweit, erscheint uns deshalb die Anonymisierung als der „Königsweg“. So stellt der Bayerische Datenschutzbeauftragte fest:

*„Die Weitergabe anonymisierter Daten oder deren Nutzung zu verschiedenen Forschungszwecken unterliegt keinen rechtlichen Beschränkungen.“*

(BayDSB, <http://www.datenschutz-bayern.de/verwaltung/epidem.htm>, 3.3).

*„Für die Anonymisierung reicht aus der Sicht des Datenschutzes die faktische Anonymisierung aus. ... Daten sind dann faktisch anonymisiert, wenn die Zuordnung zur konkreten Person nur mit unverhältnismäßigem Aufwand an Zeit, Kosten und Arbeitskraft möglich ist. Dabei dürfte es auf die Ressourcen der speichernden Stelle, also der Stelle, die mit den genannten Daten arbeiten will, ankommen.“*

(BayDSB, <http://www.datenschutz-bayern.de/verwaltung/epidem2.htm>, 2.5.1; vgl. auch §3 Abs.7 BayDSG).

Ob es sich um anonymisierte Daten handelt, bedarf also einer sorgfältigen Prüfung im Einzelfall, da abgeschätzt werden muss, ob zur De-Anonymisierung der Aufwand unverhältnismäßig hoch bzw. die Ressourcen der Projektzentrale zu knapp sind.

### 3.2 Deskriptiver Vergleich ohne statistische Adjustierung

In der BADO werden sowohl anamnestische Daten der PatientInnen, Prozessdaten der jeweiligen Behandlung, sowie Outcome-Variablen bei der Entlassung erhoben. Dieser Gliederung folgt auch die nachfolgende Datendeskription:

#### 3.2.1 Demographische und anamnestische Daten der PatientInnen

Die Kapazität der Krankenhäuser ist sehr unterschiedlich. Das größte Haus behandelte im Bezugsjahr mehr als 5mal so viele Patienten wie die kleinste Klinik. Der Anteil männlicher Patienten schwankte zwischen minimal 47% und maximal 65% (Durchschnitt: 57,5%).

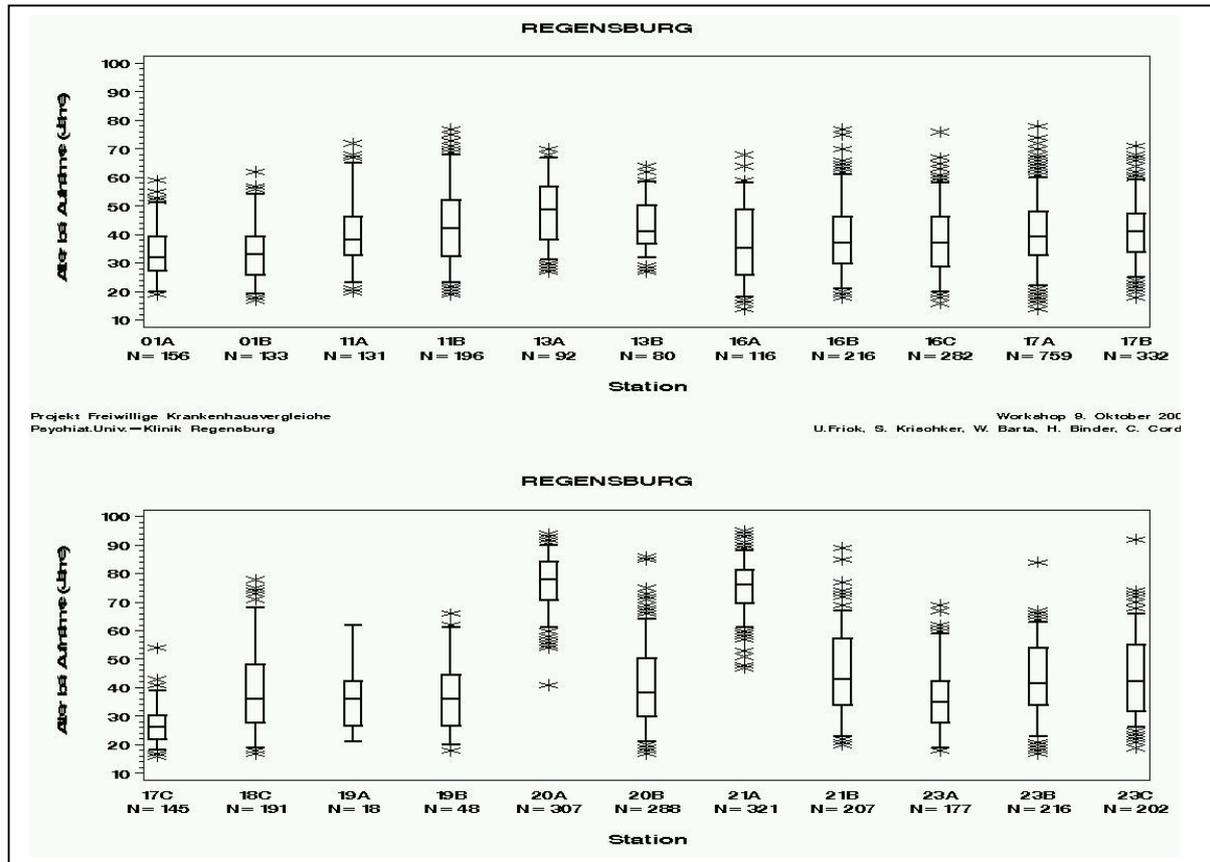
		Klinik									
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All	
Geschlecht: Männlich	nein	N	2180	1851	1043	627	1647	666	2782	1160	11956
		%	47.3	43.9	49.3	46.7	35.6	52.6	42.1	34.9	42.5
ja		N	2433	2365	1073	715	2985	601	3819	2165	16156
		%	52.7	56.1	50.7	53.3	64.4	47.4	57.9	65.1	57.5
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Die Alterstruktur der Patienten in den acht Kliniken zeigte sich als durchaus vergleichbar:

		KLINIK								
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All
Alter bei Aufnahme (Jahre)	N	4613	4216	2116	1338	4632	1267	6601	3325	28108
	Mean	44.68	45.66	46.29	45.53	41.70	43.44	46.07	41.94	44.45
	Median	41.00	41.00	42.00	43.00	39.00	40.00	42.00	38.00	40.00
	Min	14.00	13.00	16.00	8.00	13.00	16.00	13.00	15.00	8.00
	Max	95.00	94.00	94.00	94.00	96.00	92.00	99.00	97.00	99.00
	Std	17.40	18.69	18.32	18.03	17.08	16.27	18.33	16.37	17.79

Die Unterschiede zwischen den Stationen innerhalb einer Klinik sind in vielen Fällen wegen der Haus-internen Differenzierung der Behandlung beim Alter oft größer als die Unterschiede zwischen den Kliniken. Dies zeigt auch die Betrachtung des nachfolgenden Box-Plots, in dem beispielhaft die Altersverteilung in den Stationen von Regensburg angetragen wurde.

**Altersverteilung innerhalb einer einzelnen Klinik**



Die soziale Schicht, aus der die PatientInnen entstammen, kann indirekt (und mit einigen Unschärfen behaftet) aus der Schulbildung der PatientInnen abgeleitet werden. Hier zeigt sich, dass in Klinik D deutlich häufiger PatientInnen mit einer über die elementare Schulbildung hinausreichenden Ausbildung behandelt werden (45%), während das im Schnitt aller Häuser nur bei 23% der PatientInnen zutrifft. Dieser Unterschied hängt mit den abweichenden Schulsystemen der BRD und der ehemaligen DDR zusammen.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Bildung > elementar?	nein	N	3545	3448	1702	738	3618	903	5067	2644	21665
	%		76.8	81.8	80.4	55.0	78.1	71.3	76.8	79.5	77.1
ja	N	1068	768	414	604	1014	364	1534	681	6447	
	%		23.2	18.2	19.6	45.0	21.9	28.7	23.2	20.5	22.9
All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112	
	%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Hinweise auf Unterschiede im Klientel der einzelnen Kliniken, was den Schweregrad der Erkrankungen angeht, können aus den Eingangs-GAF-Werten abgelesen werden.

		KLINIK								
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All
GAF Score bei Aufnahme	N	4613	4216	2116	1340	4632	1267	6601	3325	28110
	Mean	41.57	41.44	42.02	49.51	42.03	41.24	36.27	42.39	40.88
	Median	40.00	42.00	42.00	50.00	40.00	40.16	35.00	41.33	40.00
	Min	1.00	1.00	2.00	1.00	1.00	3.00	1.00	10.00	1.00
	Max	90.00	90.00	90.00	100.00	99.00	85.00	100.00	95.00	100.00
	Std	15.51	13.89	15.53	13.57	13.82	11.09	15.28	10.11	14.44

Die PatientInnen mit dem niedrigsten Eingangs-Funktionslevel sind in der Klinik G behandelt worden. In Klinik D zeigten die PatientInnen noch das höchste Funktionslevel. Dies korrespondiert mit der Tatsache, dass von den PatientInnen aus Klinik D rund 38% vor der Hospitalisierung mit dem Partner zusammen lebten (bei einem Durchschnittswert von nur 28%, vgl. Tabelle unten). Es bleibt zu beachten, dass in Klinik B und Klinik G der Anteil von PatientInnen mit hirnanorganischen Störungen (vgl. Tabelle unten) jeweils deutlich über dem Durchschnitt liegt.

		Klinik									
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All	
Zusammenleben mit Partner											
	nein	N	3153	2881	1368	837	3600	849	4762	2787	20237
		%	68.4	68.3	64.7	62.4	77.7	67.0	72.1	83.8	72.0
ja		N	1460	1335	748	505	1032	418	1839	538	7875
		%	31.6	31.7	35.3	37.6	22.3	33.0	27.9	16.2	28.0
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Nur 16% der PatientInnen haben vor der Klinikaufnahme in einem festen Arbeitsverhältnis gestanden. Dieser Wert ist leicht höher in Klinik D (20%) und besonders hoch in Klinik C (25%). Klinik D als Universitätsklinikum und Klinik C als sehr "junge" Klinik zeigen damit ein Muster, das von dem der übrigen Versorgungskliniken abweicht. Andeutungsweise spiegelt sich das für Westdeutschland beobachtbare Nord-Süd-Gefälle in der Arbeitsmarktsituation auch bei den Beschäftigungsverhältnissen unter den stationären PsychiatriepatientInnen wider (ohne dass hier die Anonymisierung aufgehoben wäre).

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Arbeitsverhältnis vor Aufnahme	nein	N	3740	3508	1581	1073	4062	1051	5546	2987	23548
		%	81.1	83.2	74.7	80.0	87.7	83.0	84.0	89.8	83.8
ja	ja	N	873	708	535	269	570	216	1055	338	4564
		%	18.9	16.8	25.3	20.0	12.3	17.0	16.0	10.2	16.2
All	All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Das Spektrum der behandelten Diagnosengruppen wies für die schizophrenen Erkrankungsformen den höchsten Wert in Klinik F auf (41%), während in Klinik C nur rund 16% der PatientInnen diesem Erkrankungskreis zuzuordnen waren. Im Durchschnitt aller Kliniken waren 23% zu den schizophrenen Erkrankungen zu rechnen. Der Vergleichswert der bundesdeutschen Diagnosenstatistik (berechnet aufgrund der Fallzahlen im Informationssystem des Bundes: <http://www.gbe-bund.de/gbe/owa/>) betrug für schizophrene Psychosen 14,45% an allen psychiatrischen Diagnosen in 1998, wobei zu beachten ist, dass längst nicht alle psychiatrischen Erkrankungen in Psychiatrien behandelt werden. In der Klinik mit den meisten Schizophrenieerkrankungen lag damit deren Anteil um das Zweieinhalbfache höher als in der Klinik mit dem niedrigsten Anteil schizophrener Diagnosen.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Schizophrenie	nein	N	3419	3302	1784	1087	3781	744	4795	2684	21596
		%	74.1	78.3	84.3	81.0	81.6	58.7	72.6	80.7	76.8
ja	ja	N	1194	914	332	255	851	523	1806	641	6516
		%	25.9	21.7	15.7	19.0	18.4	41.3	27.4	19.3	23.2
All	All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Die affektiven Psychosen bildeten rund 8,5% der Behandlungen in den acht Vergleichskliniken. Den höchsten Anteil wies Klinik C aus mit über 18%, während in Klinik H nur 2,7% der PatientInnen zu dieser Diagnosengruppe zählte. Das heißt, der Depressionsanteil lag in der hier führenden Klinik um mehr als das 6fache höher als in der Klinik mit dem geringsten Anteil an affektiven Erkrankungen. Dies zeigt deutliche Unterschiede im Klientel der acht Häuser.

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Affektive Psychose	nein	N	3974	3842	1733	1127	4429	1059	6358	3235	25757
		%	86.1	91.1	81.9	84.0	95.6	83.6	96.3	97.3	91.6
ja		N	639	374	383	215	203	208	243	90	2355
		%	13.9	8.9	18.1	16.0	4.4	16.4	3.7	2.7	8.4
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Alkoholbedingte Erkrankungen waren am prominentesten in der Klinik D (30%), und am seltensten in Klinik F (5%) behandelt worden, bei einem Durchschnittswert von 21% alkoholbedingter Erkrankungen. Damit ist der Diagnosenanteil in der am meisten auf Alkohol spezialisierten Klinik um das 6fache höher als in der Klinik, die mit Alkoholerkrankungen am seltensten zu tun hat.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Alkoholab-hängigkeit/ missbrauch	nein	N	3544	3287	1697	939	3488	1203	5251	2742	22151
		%	76.8	78.0	80.2	70.0	75.3	94.9	79.5	82.5	78.8
ja		N	1069	929	419	403	1144	64	1350	583	5961
		%	23.2	22.0	19.8	30.0	24.7	5.1	20.5	17.5	21.2
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Auch für die Behandlung von Suchterkrankungen bei illegalen Drogen und/oder Medikamenten existieren deutliche Unterschiede zwischen den acht Vergleichskliniken. Die beiden führenden Häuser (Klinik G und Klinik H) rekrutieren bis zu einem Viertel ihres Klientels aus diesen Erkrankungen. Dagegen werden in Klinik C und F weniger als 5% der dort behandelten PatientInnen wegen Drogen- / Medikamentenproblemen aufgenommen. Im Schnitt aller Häuser sind es rund 15% des Klientels, die dem Bereich Illegale Drogen bzw. Medikamentenabusus zuzuordnen sind.

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Medikamenten-abhängigkeit/ illegale Drogen	nein	N	4325	3689	2029	1247	3458	1208	5386	2551	23893
		%	93.8	87.5	95.9	92.9	74.7	95.3	81.6	76.7	85.0
ja		N	288	527	87	95	1174	59	1215	774	4219
		%	6.2	12.5	4.1	7.1	25.3	4.7	18.4	23.3	15.0
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

PatientInnen mit neurotischen Störungen als Hauptdiagnose bildeten rund 12% der untersuchten Stichprobe. Überdurchschnittlich viele Neurosen wurden in Klinik H angegeben (21%), in Klinik C (ebenfalls 21%) und in Klinik F (19%). Dagegen berichten die Kliniken A (10%), das Haus E (9%) und die Klinik G (7%) deutlich niedrigere Werte.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Neurosen	nein	N	4163	3702	1665	1176	4223	1024	6133	2627	24713
		%	90.2	87.8	78.7	87.6	91.2	80.8	92.9	79.0	87.9
ja		N	450	514	451	166	409	243	468	698	3399
		%	9.8	12.2	21.3	12.4	8.8	19.2	7.1	21.0	12.1
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Persönlichkeitsstörung	nein	N	4305	4122	1974	1313	4373	1208	6348	3156	26799
		%	93.3	97.8	93.3	97.8	94.4	95.3	96.2	94.9	95.3
ja		N	308	94	142	29	259	59	253	169	1313
		%	6.7	2.2	6.7	2.2	5.6	4.7	3.8	5.1	4.7
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Sehr viel seltener werden bei den PatientInnen Persönlichkeitsstörungen als Hauptdiagnose registriert: im Schnitt nur bei knapp 5% der Behandlungsepisoden. Die Variation dieses Anteils zwischen den Häusern ist relativ klein. Klinik D (Minimum) berichtet von 2,2% Persönlichkeitsstörungen, Klinik A (Maximum) von deren 6,7%.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Hirnorganische Störung	nein	N	4161	3594	1892	1211	4115	1194	5777	3076	25020
		%	90.2	85.2	89.4	90.2	88.8	94.2	87.5	92.5	89.0
ja		N	452	622	224	131	517	73	824	249	3092
		%	9.8	14.8	10.6	9.8	11.2	5.8	12.5	7.5	11.0
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Auch hirnormale Störungen sind eine der eher kleineren (im Vergleich zu Schizophrenien, Affektiven Psychosen und Suchterkrankungen) Diagnosengruppen in den hier betrachteten Häusern: Ihr Gesamtanteil beträgt 11%, mit einer Schwankung von minimal rund 6% (Klinik F) bis maximal 15% (Klinik B).

Um eine erste stationär-psychiatrische Aufnahme handelte es sich im Durchschnitt bei rund 35% aller Behandlungen. M.a.W., für rund 2/3 aller Aufnahmen ist die beobachtete Behandlungsepisode eine Zweithospitalisierung oder eine Aufnahme in einer noch längeren Folge von stationären Behandlungen. Dies zeigt, dass psychische Erkrankungen oft chronisch verlaufen. Am höchsten war der Anteil an Erstaufnahmen in Klinik G (60%), am geringsten dagegen in Klinik E mit 18%. Es ist zu vermuten, dass zwischen beiden Versorgungsregionen deutliche Unterschiede in den komplementär-psychiatrischen Angeboten bestehen, und dass Klinik G im Wechsel mit den komplementären Angeboten eine andere Funktion übernimmt als das Klinikum E. Derartig hohe Unterschiede bei den Erstaufnahmen wären über eine regional differierende Inzidenz (= Zahl der Neuerkrankungen in einer bestimmten Population) aus dem reinen Morbiditätsgeschehen in der Versorgungspopulation nicht zu erklären.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Erstaufnahme?	nein	N	3151	2819	1433	890	3796	898	2667	2550	18204
		%	68.3	66.9	67.7	66.3	82.0	70.9	40.4	76.7	64.8
ja		N	1462	1397	683	452	836	369	3934	775	9908
		%	31.7	33.1	32.3	33.7	18.0	29.1	59.6	23.3	35.2
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Rund 20% aller stationär-psychiatrischen PatientInnen in den acht Häusern haben in ihrem Leben schon jemals einen Suizidversuch unternommen. Diese Zahl schwankt nur wenig zwischen den verschiedenen Häusern.

		Klinik								
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All
Jemals Suizidversuch?										
nein	N	3589	3477	1678	988	3806	958	5121	2829	22446
	%	77.8	82.5	79.3	73.6	82.2	75.6	77.6	85.1	79.8
ja	N	1024	739	438	354	826	309	1480	496	5666
	%	22.2	17.5	20.7	26.4	17.8	24.4	22.4	14.9	20.2
All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

### **3.2.2 Merkmale des Behandlungsprozesses in den Kliniken**

Eine Behandlung ohne den Einsatz von Psychopharmaka ist vor allem bei Alkoholkranken, die nur in einer akuten Intoxikationssituation aufgenommen werden ohne anschließende Behandlung, oder bei reinen Psychotherapiestationen häufiger zu registrieren. Insgesamt verlaufen knapp 32% aller stationären Aufnahmen ohne den Einsatz von Psychopharmaka. Spitzenreiter ist hier die Klinik H mit knapp 60%, während in Klinik B mit 17% der Anteil an Psychopharmaka-freien Aufenthalten am geringsten liegt.

		Klinik								
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All
Behandlung ohne Psychopharmaka										
nein	N	3154	3485	1429	909	3638	874	4378	1345	19212
	%	68.4	82.7	67.5	67.7	78.5	69.0	66.3	40.5	68.3
ja	N	1459	731	687	433	994	393	2223	1980	8900
	%	31.6	17.3	32.5	32.3	21.5	31.0	33.7	59.5	31.7
All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Analog zum hohen Prozentsatz von Psychopharmaka-Behandlungen in Klinik B ist natürlich auch damit zu rechnen, dass bei diesen Behandlungen auf die Patientenzahlen bezogen die häufigsten unerwünschten Arzneimittelwirkungen auftreten. Im

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Durchschnitt werden für 2,9% aller PatientInnen Komplikationen bei der Medikation berichtet. Die Rate liegt in Klinik B wie erwartet am höchsten (4,2%). Aber nicht in Klinik H (wo 60% aller PatientInnen ohne Psychopharmaka behandelt wurden), sondern überraschenderweise Klinik F mit nur 1,3% berichteter Komplikationen liegt sie am niedrigsten. Vor einer eingehenderen Interpretation dieser Zahlen als "Qualitätsindikatoren" müssten jedoch sehr viel genauer die einzelnen Medikationsanlässe und -strategien miteinander abgeglichen werden. Leider ist hierzu der Merkmalskatalog der BADO in seiner gegenwärtigen Form noch sehr unpräzise, so dass vor einer vorschnellen Interpretation gewarnt werden muss. In der zuständigen Arbeitsgruppe bei der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) wird diskutiert, ob in der nächsten Revision der DGPPN – BADO eine detailliertere Dokumentation der Medikation (inklusive einer „history“ der Abfolge / Art / Dosis / Compliance) aufgenommen werden soll.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Unerwünschte Arzneimittelwirkung	nein	N	4481	4039	2038	1311	4544	1250	6366	3276	27305
		%	97.1	95.8	96.3	97.7	98.1	98.7	96.4	98.5	97.1
ja		N	132	177	78	31	88	17	235	49	807
		%	2.9	4.2	3.7	2.3	1.9	1.3	3.6	1.5	2.9
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Für rund 17% der zu entlassenden Patienten werden im Durchschnitt der 8 Kliniken Möglichkeiten der Weiterversorgung in einer Komplementäreinrichtung arrangiert (ohne dass in der BADO dokumentierbar wäre, dass alle vorgeschlagenen PatientInnen auch tatsächlich in den Komplementäreinrichtungen "ankommen"). Diese "Vermittlungsquote" ist in Klinik E mit über 27% deutlich höher. Nach unten weicht die Klinik C mit 12% vom Durchschnitt um rund 5% ab.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
komplementäre Weiterversorgung?	nein	N	3895	3529	1861	1090	3368	1156	5581	2996	23476
		%	84.4	83.7	87.9	81.2	72.7	91.2	84.5	90.1	83.5
ja		N	718	687	255	252	1264	111	1020	329	4636
		%	15.6	16.3	12.1	18.8	27.3	8.8	15.5	9.9	16.5
All		N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Eine ambulante Weiterbetreuung in der hauseigenen Klinikambulanz ist natürlich nur in dem Ausmaß möglich, wie das jeweilige Haus über eine solche Ambulanz verfügt, und in dem Ausmaß, wie sie mit Therapieplätzen ausgestattet ist. Den Spitzenreiter bildet in dieser Hinsicht das Klinikum F, wo für knapp 62% der PatientInnen eine vereinbarte ambulante Weiterbetreuung am Hause für die Zeit nach der Entlassung dokumentiert wurde. Im Durchschnitt der 8 Kliniken liegt diese Zahl nur bei einem Drittel der entlassenen PatientInnen. Deutlich unterdurchschnittlich (vgl. aber oben die Quote der Vermittlungen in komplementäre Einrichtungen!) schneidet Klinik E ab, mit einem Wert von nur 12%.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
ambulante Weiterbetreuung?	nein	N	2963	2163	1356	743	4063	488	4441	2412	18629
		%	64.2	51.3	64.1	55.4	87.7	38.5	67.3	72.5	66.3
ja	N	N	1650	2053	760	599	569	779	2160	913	9483
		%	35.8	48.7	35.9	44.6	12.3	61.5	32.7	27.5	33.7
All	N	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Schließlich bleibt auch noch die Möglichkeit, dass mit einem zur Entlassung anstehenden Patienten/in eine Weiterbetreuung durch einen niedergelassenen Arzt vereinbart wurde. Dies ist im Durchschnitt bei 44% aller PatientInnen der Fall. In Klinik F (wo eine sehr hohe Vermittlung in die eigenen Ambulanz dokumentiert wurde), liegen die Zahlen sehr viel niedriger, bei lediglich 19%. Deutlicher Spitzenreiter bei der vereinbarten Nachsorge durch einen niedergelassenen Arzt ist wiederum die Klinik E, wo andererseits die geringste ambulante Weiterbehandlung registriert wurde, und auch die höchste Vermittlungsquote bei komplementären Anschlusseinrichtungen.

		Klinik								All	
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H		
Weiterbetr. niedergel. Arzt?	nein	N	2570	2228	1094	687	1707	1026	3974	2594	15880
		%	55.7	52.8	51.7	51.2	36.9	81.0	60.2	78.0	56.5
ja	N	N	2043	1988	1022	655	2925	241	2627	731	12232
		%	44.3	47.2	48.3	48.8	63.1	19.0	39.8	22.0	43.5
All	N	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
		%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Ein spezielles Merkmal des Behandlungsgeschehens innerhalb einer Klinik ist die Dichte, mit der die dort eingeführte Basisdokumentation geführt wird. Unter den 14 im Abschnitt 4.2 als Kernvariablen definierten Datenfeldern der BADO wurde in der nachfolgenden Tabelle berechnet, wie viele Angaben im Durchschnitt noch vor jedem Imputationsversuch als fehlende Werte zu registrieren waren. Dabei erreichten die Kliniken in Klinik H (6,3 missing values), Klinik F (3,9 missing values) und Klinik C (2,1 missing values) Werte deutlich über dem Durchschnitt. Die komplettesten Dokumentationen wurden in Klinik B und Klinik A geführt.

		KLINIK								
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	All
Anzahl missing values in BADO	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
	Mean	0.86	0.69	2.06	0.91	1.13	3.88	1.42	6.30	1.88
	Median	0.00	0.00	1.00	0.00	1.00	6.00	0.00	6.00	1.00
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	Max	14.00	13.00	14.00	9.00	14.00	14.00	14.00	14.00	14.00
	Std	2.39	1.36	3.14	1.36	1.40	3.22	2.48	4.64	3.15

Als Konsequenz aus diesen deutlichen Unterschieden in der Dokumentationsdichte wurden zwei Schritte unternommen, bevor die Daten in einem multivariaten Analyseverfahren über die Häuser hinweg analysiert wurden. Diese Schritte waren notwendig, damit nicht eine bestimmte PatientInnengruppe (die schwerer zu dokumentierenden) eine geringere Chance besitzt, in den Analysedatensatz des Klinikvergleichs aufgenommen zu werden. Dies hätte einen verzerrten Vergleich zur Folge.

- 1) Fehlende Werte wurden überall dort, wo es logisch plausibel erschien, aus beim betreffenden Patienten vorhandenen Angaben zu anderen BADO-Merkmalen erschlossen, oder aus für denselben Patienten vorhandenen Zweit- oder Dritthospitalisierungen (so vorhanden und als zeitkonstant anzunehmen) übertragen. Für manche Variablen (z.B. Entlassungs-GAF) wurde auch eine statistische Vorhersage des fehlenden Wertes aufgrund eines vorab geschätzten Regressionsmodelles mit vorhandenen Prädiktorvariablen (Alter, Diagnose, Vorbehandlungen u.a.) berechnet, und der so ermittelte Wert imputiert.
- 2) Dieses Verfahren bleibt auf den endgültigen Vergleich der Kliniken nur dann ohne verzerrenden Einfluss, wenn die fehlenden Werte durch den Zensierungsmechanismus MAR (missing at random; vgl. Little & Rubin, 1987) zustande gekommen sind. Weil diese Annahme eher unplausibel erscheint, wurde zusätzlich die Zahl der fehlenden Variablen pro Patient als zusätzliche Information über die PatientInnen in die statistische Modellierung mit aufgenommen, um so eine unverzerrte Schätzung der übrigen Einflussfaktoren dadurch zu ermöglichen, dass für den potenziellen Einfluss der (mangelnden) Dokumentationsdichte rechnerisch adjustiert wurde.

### 3.2.3 Merkmale des Behandlungsergebnisses

Aus dem BADO-Merkmalkatalog lassen sich einige Merkmale des Behandlungsergebnisses direkt ablesen (z.B. Schweregrad der Erkrankung im Rating nach CGI, oder GAF-Score bei Entlassung), oder aus den aktuellen und/oder früheren Behandlungsdaten berechnen.

Wird ein Patient binnen 30 Tagen nach seiner Entlassung im selben Krankenhaus wiederum stationär aufgenommen, dann kann dies als kurzfristige Rehospitalisierung und damit eher ungünstiges Behandlungsergebnis der vorangegangenen Behandlungsepisode gewertet werden. Voraussetzung dieser Zählung ist die Existenz einer auf die jeweilige Person eindeutig zuordenbaren Identifikationsziffer (die aber auch pseudonymisiert sein kann). Wird nur ein einzelnes Beobachtungsjahr im Datensatz bearbeitet, dann kann eine über die Häuser hinweg vergleichbare Zählung erst bei Aufnahmen ab dem 1. Februar des Bezugsjahres erfolgen, weil für Aufnahmen im Januar die Beobachtungsperiode von 30 Tagen noch nicht vollständig erreicht ist. In der nachfolgenden Tabelle sind die ab Februar gezählten Wiederaufnahmen binnen 30 Tagen im Verhältnis zu allen Aufnahmen im Bezugsjahr aufgelistet.

**Rehospitalisierungsquote im Bezugsjahr (gezählt ab Februar)**

		Klinik								All
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	
ab Februar: Wiederaufnahme < 30 Tage?										
nein	N	4010	3690	1813	1202	3803	1136	5766	2738	24158
	%	86.9	87.5	85.7	89.6	82.1	89.7	87.4	82.3	85.9
ja	N	603	526	303	140	829	131	835	587	3954
	%	13.1	12.5	14.3	10.4	17.9	10.3	12.6	17.7	14.1
All	N	4613	4216	2116	1342	4632	1267	6601	3325	28112
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Der Durchschnittswert von 14,1 % Wiederaufnahmen binnen 30 Tagen liegt im Vergleich zu den von Frick, Barta et. al. (2001) für das österreichische Bundesland Salzburg veröffentlichten Zahlen deutlich geringer (dort sind Wiederaufnahmen binnen 30 Tagen um 25% im langjährigen Verlauf die Regel). Auch wenn in Rechnung gestellt werden muss, dass die 14,1% Wiederaufnahmen binnen Monatsfrist nur 11/12 des Gesamtaufkommens darstellen, und für ein Gesamtjahr demzufolge mit einer Rehospitalisierungsquote von hochgerechnet 15,4% kalkuliert werden sollte, so liegen die in der BADO dokumentierten Rehospitalisierungen dennoch deutlich niedriger. Dies kann darauf zurückzuführen sein, dass mit der BADO nur innerhalb eines Krankenhauses die Personendaten zusammengeführt werden, während in der bundesweiten österreichischen Entlassungsstatistik auch landesweite Patientenmigration bei Rehospitalisierungen mit erfasst wird. Somit bedeutete eine Krankenhaus-bezogene Betrachtung einer Rehospitalisierung eine deutliche Unterschätzung des wahren Rehospitalisierungsgeschehens.

Im hier berichteten Forschungsprojekt wurde dennoch darauf verzichtet, die Personen über die beteiligten acht Krankenhäuser hinweg zusammenzuführen: Dies hätte für die drei benachbarten Kliniken Mainkofen, Landshut und Regensburg vermutlich zu einer Erhöhung der Rehospitalisierungsrate geführt, denn es finden in den Einzugsbereichen dieser drei Kliniken Patientemigrationen in die Nachbarhäuser im beachtenswerten Umfang statt. Dagegen wären die übrigen fünf Kliniken wegen der großen geographischen Distanz untereinander von einem Personenabgleich nicht betroffen worden. Um diese ungleichen Chancen bei der Schätzung der Rehospitalisierungsraten nicht zu übernehmen, wurde die Krankenhaus-bezogene Betrachtung trotz ihrer mutmaßlichen Unterschätzung der „wahren“ Rehospitalisierungsrate zum deskriptiven Vergleich der Krankenhäuser vorgezogen.

In der Interpretation der Rehospitalisierungszahlen zwischen den Krankenhäusern zeigen sich Schwankungen im Bereich zwischen 10% und 18%. Damit sind diese deskriptiven Unterschiede bei weitem kleiner, als sie beispielsweise bei der mittleren Verweildauer aufgetreten sind (vgl. nachfolgende Tabellen). Für die statistische Analyse (Abschnitt 4 dieses Berichts) im weiteren Projektverlauf wurde daher vor einer Modellierung der Rehospitalisierungsrate die Betrachtung der Verweildauer vorgezogen:

		KLINIK								All
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	
Verweildauer (episodenbezogen)	N	4609	4216	2115	1337	4632	1266	6583	3038	27796
	Mean	43.06	45.52	36.89	32.90	28.34	47.34	42.16	36.63	39.30
	Median	26.00	28.00	26.00	20.00	18.00	28.00	24.00	21.00	23.00
	Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Max	857.00	944.00	516.00	756.00	455.00	489.00	1139.00	715.00	1139.00
	Std	59.62	61.64	40.26	46.50	34.81	61.01	58.60	49.93	53.72

50% aller PatientInnen in den acht beteiligten Kliniken blieben bis höchstens 23 Tage zur Behandlung ihrer Krankheitsepisode stationär aufgenommen. Weil unter den übrigen 50% der PatientInnen aber zahlreiche Aufenthalte mit teilweise erheblich längeren Verweildauern zu verzeichnen waren, liegt das arithmetische Mittel der Verweildauern bei den 27.796 hier dokumentierten Fällen erheblich höher, nämlich bei 39,3 Tagen. Die Dichteverteilung der individuellen Verweildauern folgt in etwa einer log-Normal-Verteilung.

Die Verweildauer wurde berechnet als die Differenz in Tagen zwischen dem Entlassungsdatum und dem Aufnahmedatum, wobei mit einer Mindestdauer von 1 Tag gezählt wurde, die auch für diejenigen PatientInnen angenommen wurde, die nicht > 24 Stunden hospitalisiert waren. Der Durchschnittswert in der Klinik mit den längsten Verweildauern liegt um 67% über dem Durchschnittswert der kürzest-behandelnden Klinik.

Sollen die Unterschiede zwischen den Kliniken statistisch analysiert werden, dann empfiehlt sich eine Transformation der individuellen Verweildauern, durch die die Unterschiede bei den Verweildauer-Ausreißern nach oben eher geringer bewertet werden, und Unterschiede bei den kürzeren Verweildauern (wo auch das Gros der

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Behandlungen stattfindet) dementsprechend stärker gewichtet wird. Als Zielvariable für die statistischen Analysen des nachfolgenden Abschnitts wurde daher die Verweildauer log-transformiert ( $\log_{10}$ ). Wie sich in der nachfolgenden Tabelle ergibt, wurde dadurch die Schiefe der Verteilungen entscheidend reduziert. Dies ist ersichtlich aus der verringerten Differenz von Median und arithmetischem Mittel der logarithmierten Verweildauern.

		KLINIK								All
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H	
Logarithmierte Verweildauer	N	4609	4216	2115	1337	4632	1266	6583	3038	27796
	Mean	2.92	3.10	2.96	2.72	2.69	2.87	3.02	2.90	2.92
	Median	3.26	3.33	3.26	3.00	2.89	3.33	3.18	3.04	3.14
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	6.75	6.85	6.25	6.63	6.12	6.19	7.04	6.57	7.04
	Std	1.50	1.34	1.33	1.42	1.29	1.69	1.34	1.30	1.38

Während ihres Klinikaufenthalts erreichen die PatientInnen im Schnitt eine Verbesserung ihres sozialen Funktionslevels um 15 Punkte in der GAF (SD 15.7). Die Kliniken F und G registrierten dabei die im Schnitt größte Verbesserung (> 18 Punkte), während in Klinik H nur durchschnittlich 11 Punkte erreicht wurden. Die Klinik G war auch diejenige, die den niedrigsten Eingangs-GAF registriert hatte, aber Klinik H und Klinik F lagen jeweils im Durchschnitt bei der Eingangs-GAF. Die Klinik mit den deutlich höchsten Eingangs-GAF-Werten (Klinik D) erzielt im Mittel ziemlich genau den allgemeinen Therapiefortschritt von 15 Punkten. Am wenigsten Verbesserung des sozialen Funktionsniveaus wird im Schnitt bei den PatientInnen in der Klinik B registriert mit gut 13 Funktionspunkten Zuwachs (vgl. Abschnitt 5 für eine Analyse).

		KLINIK							
		Klinik A	Klinik B	Klinik C	Klinik D	Klinik E	Klinik F	Klinik G	Klinik H
Besserung im GAF-Score	N	4613	4216	2116	1340	4632	1267	6601	3325
	Mean	14.50	13.22	15.44	15.42	14.80	18.68	18.13	10.94
	Median	10.00	11.00	15.00	15.00	15.00	19.20	16.14	10.07
	Min	-55.00	-49.00	-64.00	-50.00	-63.00	-31.63	-60.00	-54.00
	Max	80.00	80.00	80.00	78.00	89.00	80.00	80.00	65.00
	Std	15.53	16.08	17.80	15.04	16.77	11.21	16.12	11.51

		All
Besserung im GAF-Score	N	28110
	Mean	15.09
	Median	14.25
	Min	-64.00
	Max	89.00
	Std	15.72

## 4. Ergebnisse des Multi-Level-Vergleichs von acht Kliniken

Die im Abschnitt 3 beschriebenen statistischen Kennwerte sollen im Abschnitt 4 unter Verwendung des statistischen Modells, welches im 2. Abschnitt dieses Berichtes ausführlich erläutert wurde, genauer analysiert werden. Dabei werden folgende Leitfragen bearbeitet: Unterscheiden sich die Krankenhäuser in der durchschnittlichen Verweildauer, wenn man für den jeweiligen Patienten-Mix und die jeweilige Stationszusammensetzung rechnerisch adjustiert hat (Abschnitt 4.1)? Welche Faktoren auf Patienten-Ebene wie auf Organisations-Ebene können identifiziert werden hinsichtlich ihres Einflusses auf die Verweildauer (Abschnitt 4.2)? Welche Konsequenzen für die Qualitätssicherung ergeben sich aus diesen Erkenntnissen (Abschnitt 4.3)?

### 4.1 „Where have all those differences gone?“

Mögliche Unterschiede in der durchschnittlichen Verweildauer der beteiligten acht Krankenhäuser wurden zunächst mit traditionellen varianzanalytischen Verfahren (Winer, 1971) geprüft. Als abhängige Variable wurde dazu die individuelle Verweildauer der PatientInnen log-transformiert (vgl. Brecht & Jenke, 1989; Ruttimann & Pollack, 1996; Richter, 1999) um Ergebnisverzerrungen durch Extremwerte zu verringern und die Verteilung einer Normalverteilung anzugleichen. Potenzielle Verzerrungen durch Unterschiede im Patientenmix der acht Kliniken wurden durch regressionsanalytische Adjustierung in einem ANCOVA-Modell korrigiert.

Folgende Patientenmerkmale wurden adjustiert:

- Alter und Geschlecht der Patienten,
- Diagnose des Patienten (als verschiedene Dummy-Variablen kodiert),
- GAF-Wert (als Rating für Soziale Funktionsfähigkeit, vgl. Wittchen et al. (1991)) bei Aufnahme,
- Veränderung des GAF-Wertes während des stationären Aufenthalts,
- Auftreten unerwünschter Arzneimittelwirkungen,
- ambulante Weiterbetreuung nach Entlassung,
- Weiterbetreuung im komplementären Versorgungssektor,
- Psychopharmakamedikation (J/N),
- Wiederaufnahmestatus der gegenwärtigen Behandlung,
- Dokumentationsdichte.

Diese Variablen zeigten in Voranalysen den stärksten Einfluss auf die Verweildauer (Selektionskriterium:  $t > 6$  im hierarchischen Modell, s.u.). Dieses Vorgehen hält bei der Berechnung der Einflussgröße eines Faktors den Einfluss der übrigen Merkmale jeweils konstant. Traditionelle ANCOVA-Analysen haben aber in der vorliegenden Datenstruktur wichtige Nachteile: Die Ergebnisse sind verzerrt, weil sie die Hierarchie der Datenstruktur (Patienten in Stationen in Krankenhäusern) nicht adäquat abbilden (vgl. Frick et al., 1999; Christiansen & Morris, 1997).

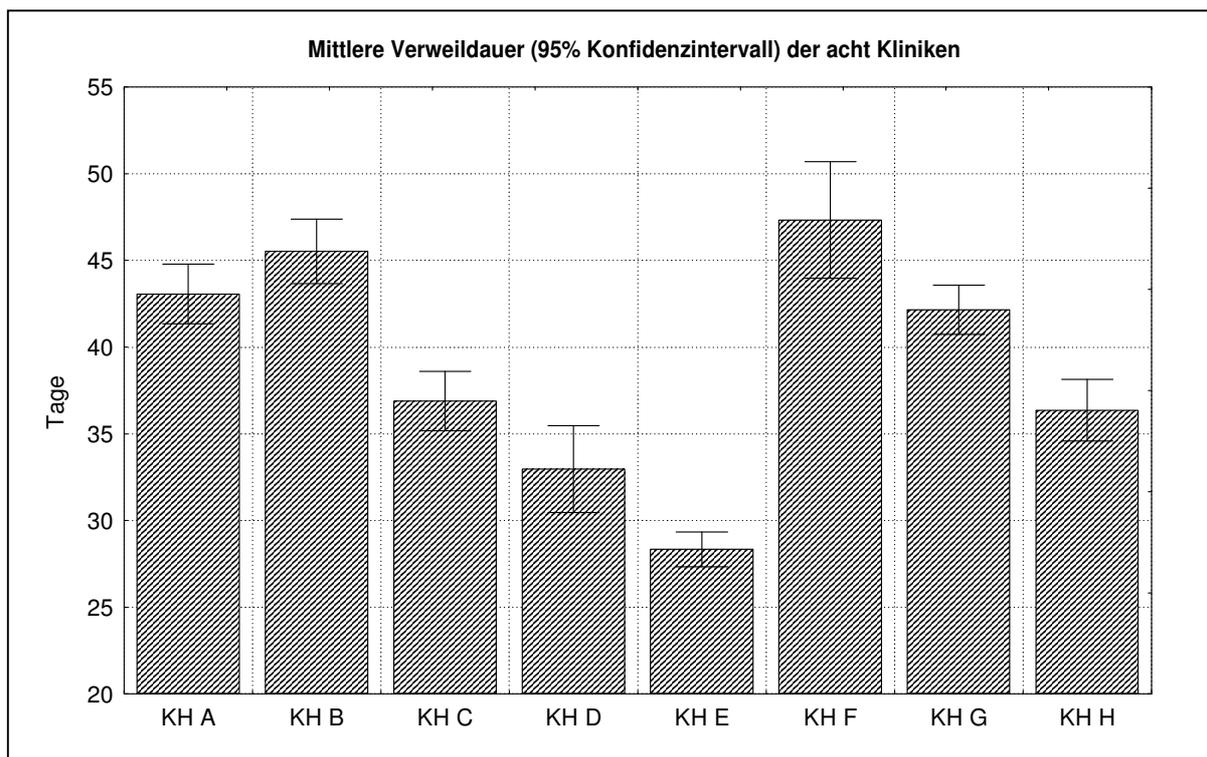
Hierarchisch-Lineare Analysen berücksichtigen die jeweiligen Datenebenen. Auf der Ebene der Patienten wurden dieselben 12 Prädiktorvariablen berücksichtigt wie im ANCOVA-Modell, und zusätzlich wurde die Datenstruktur durch Aufnahme entspre-

chender „random coefficients“ in die Schätzgleichung berücksichtigt. Eine ausführliche Darstellung von Idee und technischen Details des hier eingesetzten statistischen Modells im speziellen Zusammenhang mit Verweildaueranalysen findet sich in Abschnitt 2. Schätzprozedur und statistische Ableitung sind detailliert beschrieben bei Goldstein (1995) sowie Bryk & Raudenbush (1992). Die Durchführung der statistischen Analysen erfolgte mit dem Programm MLwiN (rel. 1.10).

Tabelle 4.1: Mittelwerte der Verweildauern in acht psychiatrischen Krankenhäusern

Klinik-Nr.	Mittelwerte der Verweildauern (Tage)	S.D.	Mittelwerte der logarithm. Verweildauer	S.D.	Adjustierte Mittelwerte (log.) <sup>&amp;</sup>
1	43.06	59.62	2.919	1.503	2.943
2	45.52	61.64	3.102	1.341	3.077
3	36.89	40.26	2.963	1.332	2.933
4	32.97	46.56	2.727	1.415	2.692
5	28.34	34.81	2.687	1.291	2.746
6	47.34	61.01	2.871	1.686	2.514
7	42.16	58.60	3.020	1.338	3.043
8	36.36	49.93	2.895	1.303	2.938

<sup>&</sup> durch lineare Regression adjustiert für 12 Patientenmerkmale



Die durchschnittliche Verweildauer in den acht Kliniken (vgl. Abschnitt 3.2.2) schwankt zwischen rund 28 Tagen und 47 Tagen. Der maximale Mittelwert liegt um 67% höher als der minimale Mittelwert.

Die Unterschiede zwischen den Kliniken deuten bei graphischer Veranschaulichung auf Überzufälligkeit hin.

Für die statistischen Analysen wurde die Verweildauer logarithmiert. Trotz des dämpfenden Einflusses auf die Unterschiede zwischen den Verweildauern zeigen sich auch hier deutliche Schwankungen zwischen den Kliniken (Minimum: 2,69; Maximum: 3,10). Damit liegt die oberste Klinik noch um gut 15% über dem Wert der minimalen Klinik (Tabelle 4.1). Techniken aus der traditionellen Varianzanalyse bestätigen diesen Eindruck: Sowohl ohne Berücksichtigung des Patientenmix (ANOVA-Modell in Tabelle 4.2), als auch mit expliziter Modellierung (ANCOVA-Modell in Tabelle 4.2) erreicht der Testwert für den Haupteffekt „Klinik-Unterschiede“ hohe statistische Signifikanz. Mit anderen Worten: auch nach statistischer Kontrolle der bedeutsamsten Patientenmerkmale unterscheiden sich die acht Kliniken hinsichtlich ihrer durchschnittlichen Verweildauer „signifikant“.

Tabelle 4.2: Varianzanalysen der logarithmierten Verweildauer

	ANOVA §	ANCOVA &
F-Test für Klinik-Effekt	38.90	49.25
df1	7	7
df2	27784	27772
p	< 0,0001	< 0,0001
erklärte Varianz der log.VWD	0.97%	21.77%

§ Varianzanalyse mit Haupteffekt Klinik

& Kovarianzanalyse mit Haupteffekt Klinik und 12 Patientenmerkmalen

Dieses Bild wandelt sich dramatisch, wenn die hierarchische Struktur der Daten korrekt modelliert wird: Unterschiede zwischen Krankenhäusern sind nahezu verschwunden ( $t < 1$ ; n.s.), wenn sowohl die Patienten-Ebene wie die Stationsebene als in den Krankenhäusern jeweils geschachtelte Datenstrukturen berücksichtigt werden (vgl. Tabelle 4.3).

Tabelle 4.3: Varianzkomponenten in der Hierarchischen Analyse

Ebene	Varianz-Schätzer <sup>&amp;</sup>	S.E.	t-Wert	p	% der Gesamt-varianz <sup>§</sup>
Krankenhaus	0.0024	0.0095	0.25	n.s.	0.00
Station	0.2692	0.0360	7.48	p < 0.0001	18.25
Patient	1.3343	0.0113	118.08	p < 0.0001	81.75

& 3 Level mit 12 Prädiktoren auf Level1 = Patient

§ basierend auf dem Random-Intercept-ANOVA Modell ohne zusätzliche Prädiktoren

Hierarchische Analysen erlauben eine Abschätzung der relativen Varianzanteile, die den Ebenen zuzuordnen sind: Zwischen den beteiligten Kliniken (Ebene 3) ergibt sich demnach nur noch ein Varianzanteil, der nicht mehr von Null unterscheidbar ist. Die Stationen (Ebene 2) binden dagegen rund 18% der Gesamtvarianz der logarithmierten Verweildauer, rund 82% sind der Ebene 1 und damit den Patienten zuzuordnen.

Diese Varianzanteile geben keinerlei Auskunft darüber, wie hoch die erklärte Varianz auf jedem untersuchten Level liegt, sondern bezeichnen lediglich, wie viel Varianz auf dem jeweiligen Level gebunden ist. Es sind also auf der Ebene der Klinik 0% der Verweildauervariation gebunden, und daher kann auch keinerlei Merkmal der Klinik zur Erklärung der unterschiedlichen Verweildauern einen Beitrag leisten. Die in der

varianzanalytischen Betrachtungsweise festgestellten „signifikanten“ Klinikunterschiede in den Verweildauern sind somit voll auf die Unterschiede in Patientenmix und Stationszusammensetzung rückführbar. Insgesamt sind 18 % der Verweildauervarianz auf der Stationsebene gebunden. Stationsmerkmale wie diagnostische Spezialisierung, bauliche Merkmale, Stationsgröße und vieles mehr könnten also maximal 18% der Verweildauervarianz aufklären. Der Löwenanteil der Varianz liegt auf der Ebene der Patienten gebunden. Dies bedeutet aber nicht, dass mit dem hier eingesetzten statistischen Modell auch 82% der Varianz aufgeklärt werden können (vgl. Frick et al., 1999, für die erklärte Varianz aufgrund von Patientenmerkmalen).

**Zusammenfassend** kann also festgehalten werden: Durch den Einsatz von Hierarchischen Analysen wurde beim Vergleich von 8 Kliniken herausgearbeitet, dass augenscheinliche Unterschiede in der Behandlungsdauer dem unterschiedlichen Patientenmix (Löwenanteil von 82% der Varianz) und der unterschiedlichen Ausgestaltung der verschiedenen Stationen (18% der Varianz) zugeordnet werden können. Aus dem Blickwinkel der gegenwärtigen Qualitätssicherungsdiskussion ergibt sich eine wichtige methodische Minimalforderung für künftige Vergleiche psychiatrischer Krankenhäuser. Reine Level-3-Vergleiche, also die Gegenüberstellung von nicht weiter adjustierten Klinikmittelwerten, sollten ebenso abgelehnt werden wie Vergleiche, die mit inadäquaten Methoden verzerrte Adjustierungen liefern können (vgl. auch DGPPN, 2001). Für die Qualitätssicherung ergibt sich daraus eine Aufforderung zu einer Refokussierung des analytischen Blickes: Es gilt, Charakteristika der behandelnden Stationen, und weniger Charakteristika der Kliniken zu untersuchen. Nicht die „Krankenhauskultur“, sondern die jeweilige „Stationskultur“ beeinflusst in unserer Stichprobe die durchschnittliche Behandlungsdauer zu 18% der Varianz.

Einer detaillierteren Beschreibung der gefundenen Einflussfaktoren auf individueller Patientenebene und der Organisationsebene der Stationen bildet daher den Schwerpunkt im nachfolgenden Abschnitt 4.2 .

## **4.2 Einflussfaktoren auf die Verweildauer**

Aus den Ergebnissen von Abschnitt 4.1 ergibt sich, dass die Merkmale, die den Krankenhäusern speziell anzurechnen sind (z.B. Bettenzahl des Krankenhauses, Verhältnis der Bettenzahl zur Größe der Versorgungsbevölkerung, Status des Krankenhauses als reine Versorgungsklinik versus als Universitätsklinik, usw.) nicht mehr geprüft werden müssen auf ihren möglichen Zusammenhang mit der in den acht Häusern erzielten Verweildauer. Da die den Krankenhäusern zuzurechnenden Varianzanteile Null sind, kann auch nicht versucht werden, auf dieser Datenebene durch Einführung von Prädiktoren nach dem Muster von Gleichung (3a) oder (4a) (vgl. Abschnitt 2, allerdings erweitert auf einen 3-Level-Ansatz) diese nicht vorhandenen Varianzanteile weiterhin durch spezielle Variablen aufzuklären.

Sollten beispielsweise die beiden Universitätskliniken unter den acht Häusern (Jena und Regensburg) deutlich kürzere Verweildauern erzielen als die übrigen Häuser<sup>2</sup>, dann wäre eine Interpretation nach dem Muster „Universitätskliniken behandeln im Durchschnitt schneller als Versorgungskrankenhäuser“ ein klassischer ökologischer Fehlschluss (vgl. zur Diskussion über diesen Terminus technicus unlängst: Babor,

---

<sup>2</sup> Dies ist nur ein Beispiel. Empirisch ist dies nicht der Fall.

2000 beim Kommentar zu Smart & Mann, 2000). Denn dieser –**und auch alle anderen** auf Datenebene 3 (Krankenhäuser) scheinbar verortbaren – Unterschied(e) sind, nachdem für den Patienten-Mix und die Stationsorganisation geeignet adjustiert wurde, vollkommen bedeutungslos.

Dieses Ergebnis bedeutet nur eine Aussage über die hier untersuchten 8 Häuser. Ob diese Aussage auch Bestand hat bei einer Untersuchung einer größeren Anzahl von Häusern, oder bei der Einbeziehung beispielsweise von psychiatrischen Abteilungen (hier wurden ausschließlich psychiatrische Krankenhäuser analysiert) in den Vergleich, kann nicht mit Sicherheit vorhergesagt werden. Auch hierarchische Analysen sind natürlich in der statistischen Power, mit der Unterschiede einer vorgegebenen Effektgröße gefunden werden können, von der Fallzahl abhängig. Durch die Einbeziehung von deutlich mehr Krankenhäusern (bei ansonsten gleichen Verhältnissen) könnten möglicherweise auch kleine Unterschiede auf der Datenebene 3 noch gesichert werden. Aber wichtiger als diese rein methodische Einschränkung erscheint doch der Hinweis, dass durch eine andere Stichproben-Zusammensetzung eben auch inhaltlich eine neue Situation möglich wird. Der freiwillige Vergleich durch die beteiligten acht Häuser weist diese Häuser gerade durch ihre Teilnahme am Modellprojekt als besondere Anstalten aus: 1. war in allen diesen Häusern die von der DGPPN empfohlene Basisdokumentation im Stichjahr bereits seit mehreren Jahren als Routedokumentation implementiert. Dies ist zwar bei zunehmend mehr Krankenhäusern, aber leider bei weitem nicht bei allen Krankenhäusern der Fall. 2. Waren die Häuser auch bereit, ihre anonymisierten Behandlungsdaten in einen gemeinsamen Datenpool einzubringen. Diese Bereitschaft hängt möglicherweise zusammen mit einer (ggf. auch via BADO empirisch begründbaren) Gewissheit, bei den eigenen Behandlungsgewohnheiten keine exzeptionell auffälligen Besonderheiten zu erzielen. Es ist anzunehmen, dass in Kombination beider Umstände (1. Existenz einer funktionierenden BADO, 2. Bereitschaft, sich selber transparent zu gestalten) eine Selektion von speziellen Krankenhäusern gefunden wurde, die sich auch in ihren erzielten Verweildauern näher sind, als es das gesamte Spektrum aller bundesdeutschen psychiatrischen Kliniken bzw. Abteilungen erreichen könnte. Es bleibt daher weiteren Untersuchungen vorbehalten, darüber zu entscheiden, ob in der bundesdeutschen psychiatrischen Versorgung die Klinik endgültig kein bedeutsamer Einflussfaktor auf die Verweildauer ist, oder ob doch für andere Abteilungen/Häuser solche unterschiedlichen „practice styles“ sicherbar sind.

Diese Argumentation entwertet aber nicht die in den acht nunmehr verglichenen Anstalten erzielten Ergebnisse. Untersucht werden können nämlich durch das Datenpooling und die eingesetzte Methodik hierarchischer Analysen nunmehr bislang noch nicht ausreichend beforschte Muster, wie organisatorische, Stations-bezogene, und individuelle, Patienten-bezogene Einflussfaktoren beim Zustandekommen der Verweildauern in psychiatrischer stationärer Behandlung in Wechselwirkung beteiligt sind.

Im Datensatz der 8 Kliniken mit ihren 135 Stationen konnte für die Stationsebene im Rahmen dieses Forschungsprojektes nicht dieselbe Anzahl an Stationsmerkmalen erhoben werden, wie sie innerhalb unseres eigenen Hauses Regensburg zur Analyse im Abschnitt 2 (vgl. oben) beigebracht werden konnte.

Im Projektverlauf wurde zwar ein dementsprechender „Stationsstrukturfragebogen“ entwickelt, der für eine ganze Liste von Stationsmerkmalen konkrete Ausstattungs-

merkmale baulicher, organisatorischer und personeller Art quantitativ erfasst. Dieser neu entwickelte Fragebogen wurde auch an die 135 Stationen versandt. Jedoch zeigte sich empirisch der Rücklauf derart unbefriedigend, dass Informationen aus dieser Befragungsaktion nicht in diese Auswertung aufgenommen werden konnten. Der mangelnde Rücklauf war dabei keineswegs in erster Linie auf eine ungenügende „Auskunftsberedtheit“ zurückzuführen. Vielmehr standen den Häusern manche Informationen nach dem Ablauf von 1 oder 2 Jahren nicht mehr zur Verfügung. Beispielsweise die Dienststellenpläne der Stationen, aus denen die Vollzeitäquivalente an therapeutischen und/oder pflegerischen Kräften auf den einzelnen Stationen im betreffenden Jahr hätten rekonstruiert werden können, waren nicht mehr bei vertretbarem Aufwand bebringbar. Als Konsequenz aus diesen Umständen lässt sich die Anregung für künftige Institutionsvergleiche ziehen, diese hinsichtlich der Erhebung von Daten aus mehreren institutionellen Quellen und auf mehreren logischen Ebenen tunlichst prospektiv anzulegen, und vom Aufwand her keinesfalls zu gering zu planen. Als Konsequenz im Rahmen des hier zu berichtenden Forschungsprojektes wurden als Stationsmerkmale nur diejenigen Merkmale in die Analyse mit einbezogen, die sich aus den Patientenmerkmalen unmittelbar durch Aggregation berechnen ließen.

Analog zum Vorgehen im Abschnitt 2 wurden auch die mehr als 27000 Datensätze der Vergleichshäuser einer Multi-Level-Analyse unterzogen. Im finalen statistischen Modell fanden sich dabei vier unterschiedliche Typen von Einflussfaktoren für die Verweildauer (Tab. 4.4). Die jeweiligen Koeffizientenschätzungen sind in der Tabelle 4.5 aufgeführt (siehe unten).

#### **4.2.1 Über die Stationen konstante Patienten-Effekte**

Diese Variablen entsprechen in der Art ihres Zusammenhangs dem „normalen“ Muster, wie es von multiplen linearen Regressionen her bekannt ist: Je mehr ein bestimmtes Patientenmerkmal zutrifft, umso länger bzw. kürzer (je nach Richtung des rechnerischen Zusammenhangs) bleibt der betreffende Patient hospitalisiert. Zu beachten bleibt bei der Interpretation, dass durch den multivariaten Ansatz nur „Netto-Effekte“ beschrieben werden. Das bedeutet, dass der mögliche Einfluss anderer Variablen schon „herausgerechnet“ wurde.

Wenn ein Patient / eine Patientin als Wiederaufnahme in der BADO registriert wurde, dann war dies mit durchschnittlich 0.20485 log-Tagen Aufenthaltsverkürzung verbunden. Umgekehrt formuliert: Erstbehandlungen dauerten in der Regel länger. Dies gilt für alle Häuser und alle Stationen.

Wurde bei der Medikation eine Komplikation verzeichnet, so waren die betreffenden Aufenthalte im Schnitt um 0.4547 log-Tage länger. Auch die Diagnose einer affektiven Störung war mit einer institutionsübergreifend einheitlichen Verlängerung der Durchschnittsverweildauer um 0.44 log-Tage verbunden. Sowohl die Vereinbarung einer an den akuten stationären Aufenthalt anschließenden Weiterversorgung im komplementären Sektor (+ 0.54 log-Tage), wie in geringerem Ausmaß auch die Übernahme in die ambulante Weiterbehandlung am eigenen Haus (+ 0.14 log-Tage) waren verbunden mit einer zeitlichen Verlängerung der betrachteten Behandlungsepisode.

Schließlich führte jeder GAF-Punkt, um den sich ein Patient / eine Patientin im Laufe ihres stationären Aufenthalts in seinem / ihrem sozialen Funktionsniveau verbesserte, zu einem um durchschnittlich 0.01 log-Tage längeren Aufenthalt. Dies gilt unabhängig vom GAF-Eingangsniveau, das der Patient / die Patientin mit in die Klinik brachte, weil für den (insignifikanten) Einfluss des Eingangs-GAF-Scores im Modell aus theoretischen Gründen rechnerisch adjustiert wurde (nicht in Tab. 4.4 aufgeführt, aber vgl. Tab.4.5). Es ist beachtenswert, dass dieser „therapeutische Investitions-Score“ über alle Häuser und Stationen hinweg als einheitlicher Effekt berechnet werden konnte!

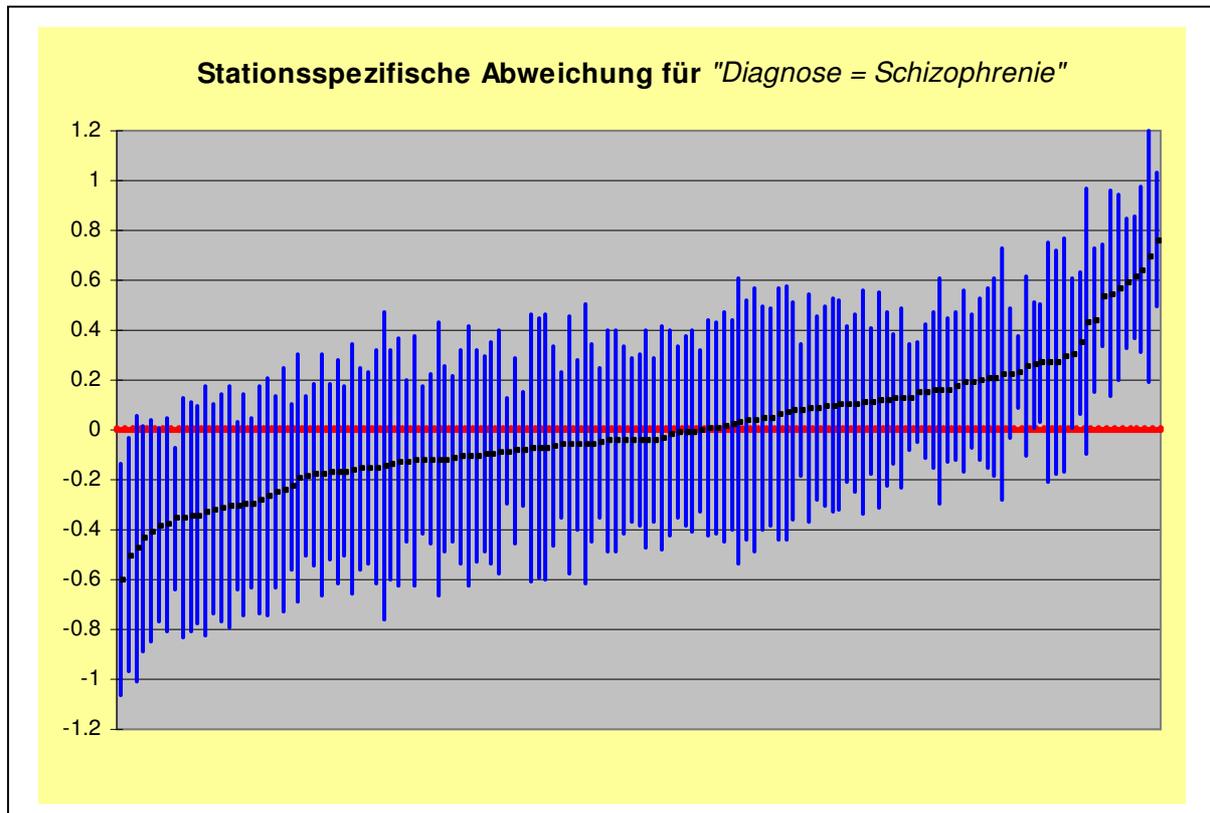
<b>Tab. 4.4 Einflussgrößen auf Patienten-Ebene</b>			
<b>Über die Stationen konstante Effekte</b>		<b>Über die Stationen variierende Effekte</b>	
<i>Variable</i>	<i>Richtung</i>	<i>Variable</i>	<i>Richtung</i>
Wiederaufnahme?	-	Alter bei Aufnahme	+
Medikationskomplikation?	+	Diagnose: schizophrene Psychose?	+
Weiterversorgung im komplementären Bereich?	+	Behandlung ohne Psychopharmaka?	-
Diagnose: Affektive Psychose?	+	Anzahl der missing values	+
Ambulante Weiterbetreuung?	+		
Verbesserung im GAF-Score	+		
<b>Einflussgrößen auf Stations-Ebene</b>			
<b>Stationseffekte ohne Wechselwirkung mit Patienten-Merkmalen</b>		<b>Stationseffekte in Wechselwirkung mit Patienten-Merkmalen („cross-level interaction“)</b>	
<i>Variable</i>	<i>Richtung</i>	<i>Variable</i>	
Zahl der Patienten / anno	-	Slope effect des Jahresanteils schizophrener Patienten auf den Einfluss der Diagnose schizophrene Psychose	
Jahresanteil der Patienten mit neurotischen Störungen	+	Slope effect des Jahresanteils an Patienten ohne Psychopharmaka auf den Effekt der Pharmaka-freien Behandlung	
Stationsbezogener Bonus/Malus oder „Intercept-Effekt“ („practice style“)			

Alle anderen im finalen statistischen Modell geschätzten Effekte für die Verweildauer können nicht so einfach und unmittelbar in ihrer Wirkung beschrieben werden: Sie hängen jeweils davon ab, welche der 135 beteiligten Stationen gerade betrachtet wird. Daher wird in den nachfolgenden Darstellungen jeweils der durchschnittliche Zusammenhang zwischen Patientenmerkmal und Verweildauer berichtet, aber auch graphisch veranschaulicht, wie stark sich die jeweils extremsten der 135 Stationen in ihrer Überformung auf diesen Durchschnittseffekt ausgewirkt haben. Die Benennungen der Stationen bleiben dabei im Rahmen dieses Berichts anonym, wurden den Verantwortlichen aus den 8 Häusern jedoch de-anonymisiert mitgeteilt.

#### **4.2.2 Über die Stationen variierende Effekte**

Wenn ein(e) Patient(in) mit einer Diagnose aus dem schizophrenen Formenkreis stationär aufgenommen wurde, dann wurde im Schnitt eine um 0.19 log-Tage längere stationäre Behandlung erbracht als bei den übrigen PatientInnen. Je nach Station,

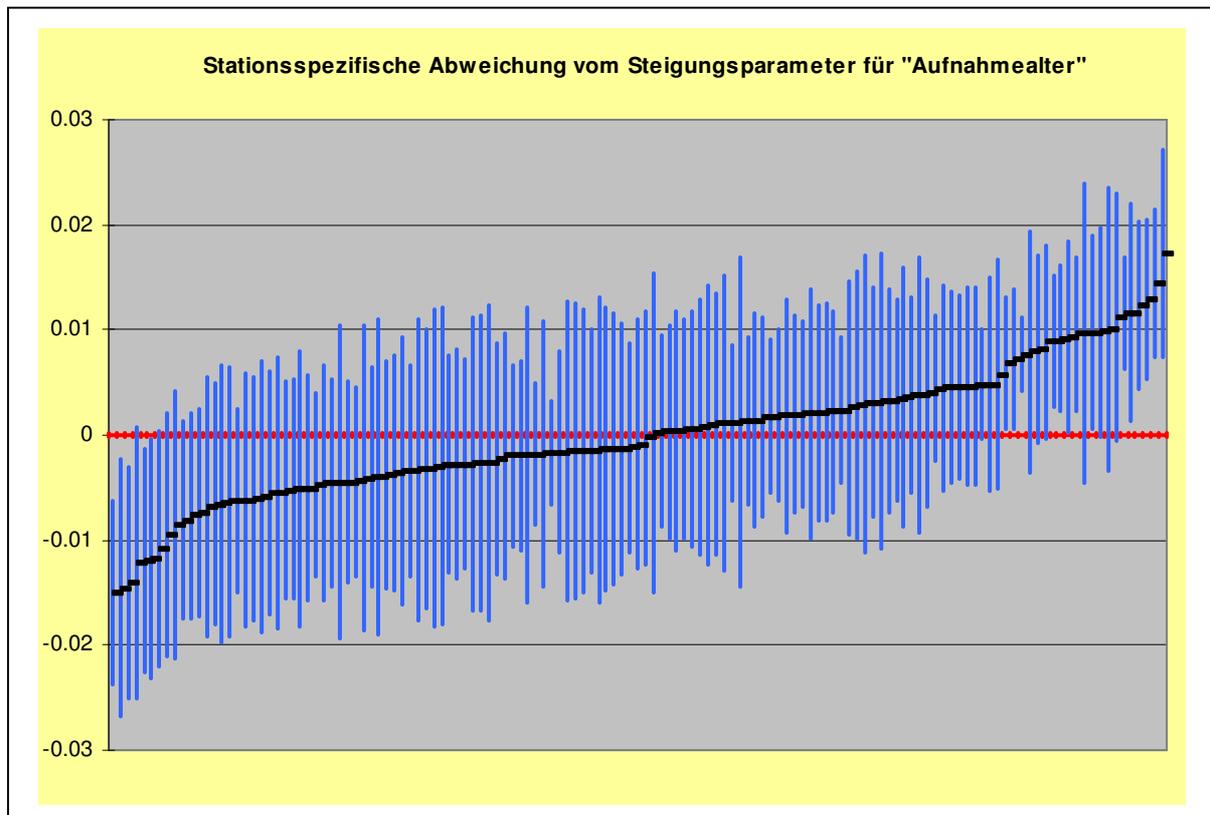
aus der der/die Patient(in) entlassen wurde, sind zu diesem durchschnittlichen „Diagnosen-Malus“ noch stationsspezifische Zuschläge bis 0.76 log-Tage, bzw. Abschläge bis zum Ausmaß von  $-0.60$  log-Tage hinzuzurechnen. Nachfolgende Abbildung visualisiert dies:



Zusätzlich zur Schätzung des stationsspezifischen Zu- bzw. Abschlags auf den „Schizophrenie-Malus“ bei der Verweildauer ist im obigen „Raugen-Plot“ auch noch angetragen, wie sicher dieser stationsspezifische Term berechnet werden konnte. Die Breite des jeweiligen Konfidenzintervalls ergibt sich aus der Varianz der Verweildauern in der jeweiligen Station und aus den Fallzahlen im Betrachtungsjahr. Diejenigen Stationen, bei denen das Vertrauensintervall nicht die Null-Linie (kein Zu- oder Abschlag auf den Durchschnitts-Malus von 0.19 log-Tagen) überstreicht, können im statistischen Sinne als „auffällige“ Stationen bezeichnet werden. In diesen Stationen wird mit der Bedeutung, die eine Diagnose vom schizophrenen Formenkreis für die Verweildauer besitzt, signifikant anders (stärkerer oder schwächerer Verlängerungseffekt) umgegangen, als dies beim Durchschnitt der 135 Stationen insgesamt der Fall ist. Dies trifft für insgesamt 17 Stationen zu: 3 Stationen behandeln „Schizophrenie“ sogar von der Tendenz her als Verweildauer-verkürzendes Merkmal, 14 Stationen liegen deutlich über dem Stationsdurchschnitt, weil sie bei dieser Diagnose eine ungewöhnlich verlängerte Verweildauer erzielt haben.

Für das Lebensalter der Patienten (bei ihrer Aufnahme) gilt: Für jedes Lebensjahr verlängert sich die Verweildauer um durchschnittlich 0.0036 log-Tage (vgl. Tab. 4.2.2). Allerdings ist auch hier wieder von zentraler Bedeutung, auf welcher Station die Entlassungsentscheidung gefällt wurde. Die Station, die das Lebensalter am stärksten Verweildauer-verlängernd berücksichtigt hat, erzielte eine Abweichung vom Durchschnittswert um 0.017 log-Tage pro Lebensjahr nach oben. Am Gegenpol konnte eine Station gefunden werden, die das Lebensalter sogar eher Verweildauer-

verkürzend berücksichtigt hat in ihren Entlassungsentscheidungen, mit einer stationsspezifischen Abweichung von  $-0.015$  log-Tagen pro Lebensjahr (vom Stationsdurchschnitt beim Lebensalter-Effekt auf die Verweildauer).

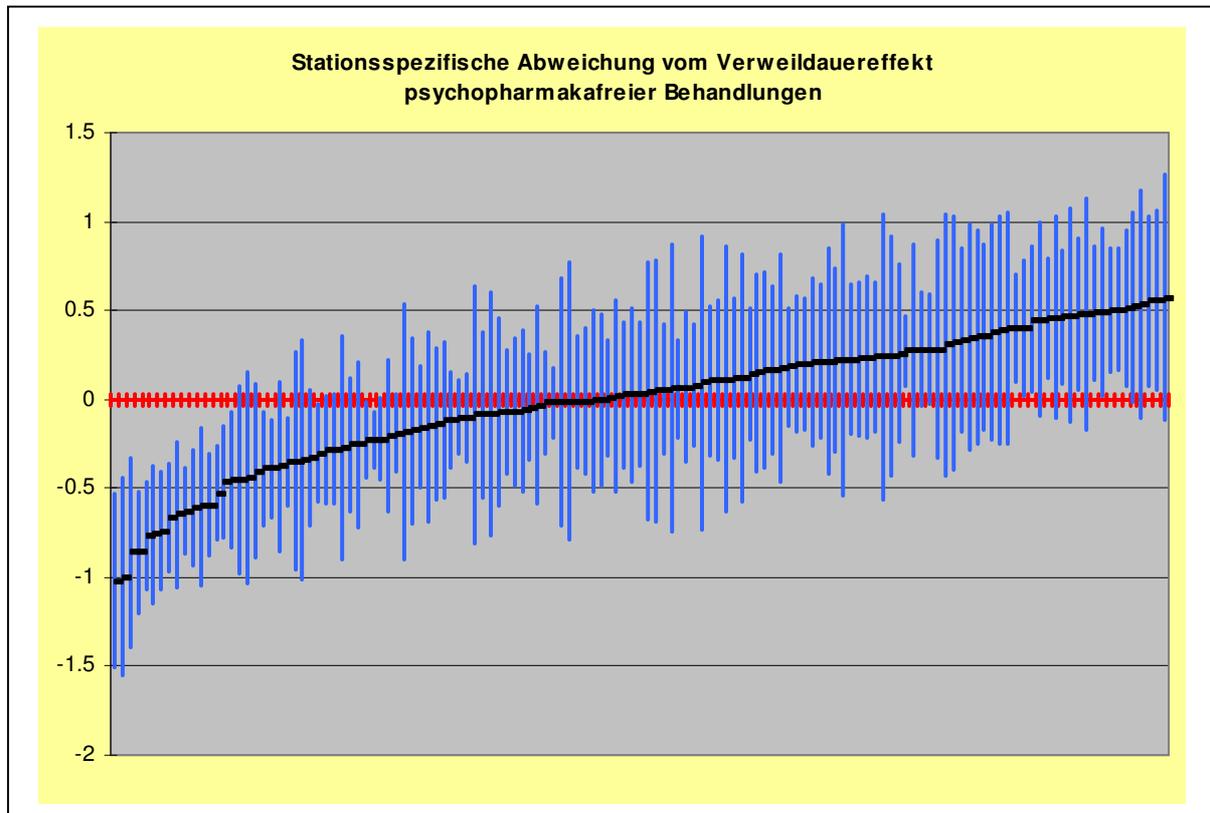


Fünf Stationen waren statistisch auffällige „Alters-Unterbewerter“. Für 13 Stationen galt, dass sie das Lebensalter weitaus dramatischer als die übrigen Stationen als Verweildauer-verlängernd gewertet haben in ihren Entlassungsentscheidungen.

Die Tatsache, dass ein Patient / eine Patientin während der beobachteten Behandlungsepisode behandelt wurde ohne den Einsatz von Psychopharmaka, stand in einem sehr deutlich über die Stationen hin wechselnden Konnex mit der Verweildauer. Im Durchschnitt aller Stationen war der Nichteinsatz von Psychopharmaka mit einer sehr deutlichen Verkürzung der Verweildauer ( $-0.96$  log-Tage) verbunden. Insgesamt 20 Stationen lagen in ihrer Verweildauer-Verkürzung bei fehlenden Psychopharmaka sogar noch signifikant unter diesem Durchschnittswert. Die Station mit dem größten „Abschlag“ auf den sowieso schon im Durchschnitt als „Abschlag“ wirksamen Koeffizienten erreichte eine zusätzliche, stationsspezifische Verringerung der Verweildauer um  $-1.02$  log-Tage. Stationen, die eine derart massive Verkürzung der Verweildauer vornehmen, tun dies (wie sich bei einem Workshop der beteiligten Krankenhäuser zur Interpretation dieser Daten herausstellte) meist, weil es sich bei den solchermaßen behandelten PatientInnen um Alkoholikranke handelt, die ohne weitere medizinische Maßnahme nur für 1 Nacht in einer akuten Vergiftung aufgenommen worden waren.

Es gibt aber noch einen zweiten Grund, warum in Stationen das Patientenmerkmal „Behandlung ohne Psychopharmaka“ eine deutliche Abweichung von seinem Durchschnittseffekt (Verkürzung um  $0.96$  log-Tage) in der entgegengesetzten Richtung aufweist. Insgesamt 14 Stationen weichen statistisch signifikant nach oben ab, d.h.

sie behandeln PatientInnen ohne den Einsatz von Psychopharmaka eher länger als die übrigen Stationen. Diese Abweichung nach oben geht bis zu einer maximalen Effektgröße von +0.57 log-Tagen. Bei den hier anzuführenden Stationen handelt es sich meist um Psychotherapie-Stationen.



Neben diesen drei inhaltlich zu interpretierenden, stationsspezifisch überformten Effekten von Patientenmerkmalen auf die erzielte Verweildauer zeigte sich im finalen statistischen Modell auch eine aus methodischen Gründen eingeführte Variable als überzufällig bedeutsam. Die Anzahl von fehlenden Werten (gezählt von 0 bis 14 bei einer Liste von vorab definierten Kernvariablen der BADO, nicht deckungsgleich mit den letztlich hier in der Analyse verwendeten Prädiktoren) wurde als Merkmal gesondert erhoben. Auf eine inhaltliche Interpretation dieses Merkmals wird jedoch im Rahmen dieses Berichtes verzichtet, weil ja ein Großteil der Werte aus „Kollateralinformation“ erschlossen werden konnte, bzw. aus anderen Variablen statistisch geschätzt wurde, und somit in die endgültige Analyse nicht als „missing“ einging. Zur Vermeidung von Selektionsartefakten bei informativem Zensierungsprozess wurde die Variable „Anzahl der missing values“ als Patientenmerkmal erhoben, und in, – wie erwartet, – signifikanter stationsspezifischer Variation als Random Koeffizient im Vorhersage-Modell berücksichtigt. Es hängt also sowohl von den Patientenmerkmalen, wie von der Station ab, ob bei besonders lang verweilenden PatientInnen die Dokumentationsdichte eher besser, oder eher schlechter (Mehrzahl der Fälle) ist. Im Rahmen dieses Berichtes wird die „Anzahl der missing values“ ausschließlich als konfundierender Faktor betrachtet, für den die übrigen, inhaltlich bedeutsamen Variablen adjustiert werden.

### **4.2.3 Stationseffekte ohne Wechselwirkung mit Patienten-Merkmalen**

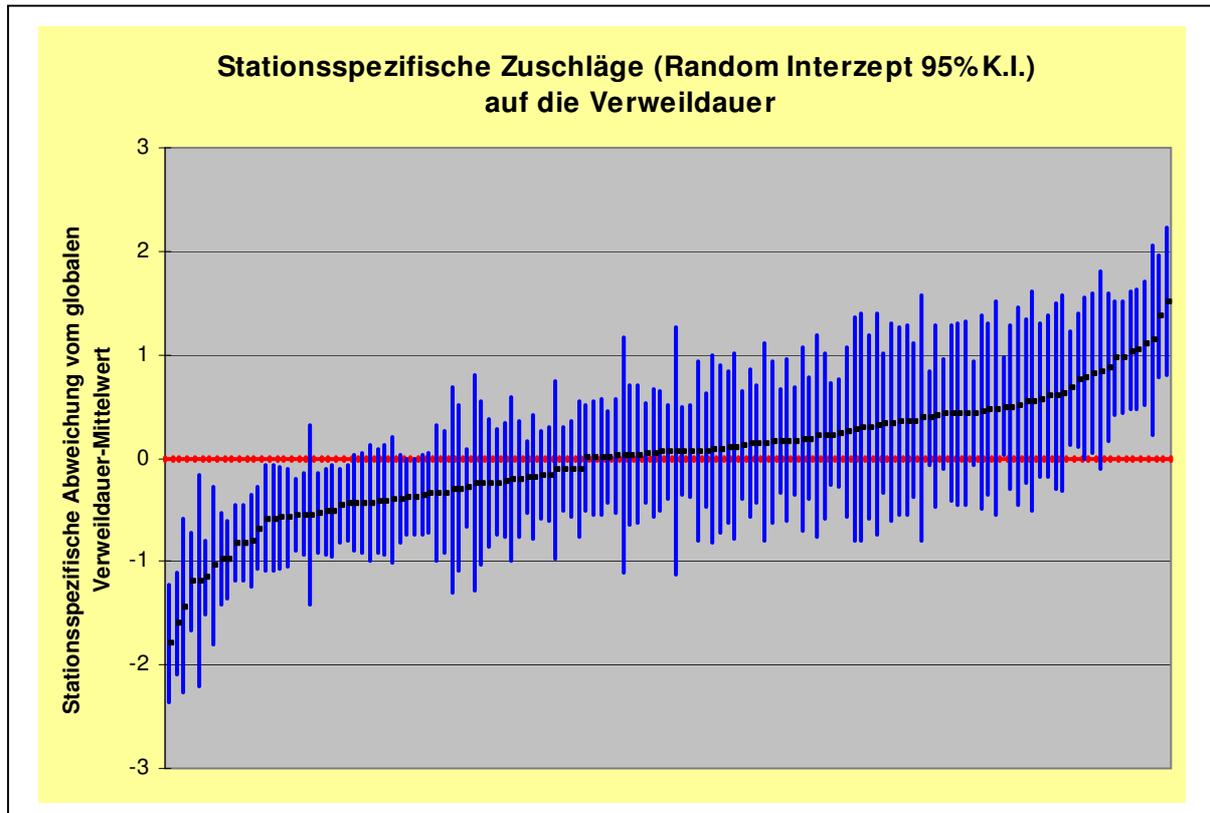
Aus den BADO-Daten können verschiedene Indikatoren für die Größe und das Leistungsprofil einer Station berechnet werden. Als Indikator für die Größe einer Station, aber auch als Maß für die Arbeitsbelastung des Pflegepersonals wurde der jährliche „Patientendurchsatz“ als Gesamtzahl aller im Bezugsjahr aus der Station entlassenen PatientInnen berechnet. Je mehr PatientInnen die betreffende Station in einem Jahr aus der Behandlung nach außerhalb der Klinik entlässt, umso kürzer werden tendenziell die dort erzielten Verweildauern (um -0.00023 log-Tage kürzere Durchschnittsverweildauer pro zusätzlichem Jahrespatienten vgl. Tab.4.5).

Dies kann einerseits gedeutet werden als eine Reaktion auf einen hohen Aufnahmepressur von außen, der die Verkürzung von Verweildauern zur Wahrung einer Mindestversorgung möglichst vieler PatientInnen erzwingt. Andererseits lässt sich die kausale Richtung dieses Zusammenhangs allerdings auch genau umgekehrt lesen: Wenn PatientInnen in einer Station besonders kurz behandelt werden, dann wird zur Wahrung der Auslastungsquote der nächste Patient / die nächste Patientin möglichst sofort wieder aufgenommen (aus allen beteiligten Krankenhäusern war ein hoher „Aufnahmepressur“ aus der Versorgungspopulation berichtet worden). Dies führt dann dazu, dass die Zahl der innerhalb eines Jahres behandelten PatientInnen steigt.

Je mehr PatientInnen mit neurotischen Störungen innerhalb einer Station behandelt werden, umso länger dauert die Behandlung für alle dort verweilenden PatientInnen. Die Interpretation dieses Level-2 Effekts bedarf besonderer Sorgfalt: Es handelt sich nämlich nicht um einen Patienten-bezogenen Effekt. Nicht die Tatsache, dass ein bestimmter Patient wegen einer neurotischen Störung stationär behandelt wird, lässt eine längere Verweildauer erwarten. Ein derartiger Effekt hätte sich in der statistischen Analyse als Level-1 Effekt (vgl. oben, Abschnitt 4.2.1) durchsetzen müssen, und wurde auch geprüft und als bedeutungslos verworfen. Es handelt sich hier um einen Kontexteffekt in dem Sinne, dass in Stationen mit sehr vielen neurotischen Patienten (gemessen als der Jahresanteil der Neurotiker am gesamten Patientenaufkommen der jeweiligen Station) für alle dort behandelten Patienten (also sowohl die Neurotiker wie alle anderen) längere Verweildauern beschließen, als nach den individuellen Merkmalen der PatientInnen jeweils zu erwarten gewesen wäre. Unter Qualitätssicherungsgesichtspunkten bleibt zu fragen, ob die gemeinsame Behandlung von neurotischen und nicht-neurotischen PatientInnen für letztere damit nicht einen Behandlungsnachteil mit sich bringt? Bevor allerdings vorschnelle Interventionen (z.B. andere Spezialisierungen von Stationen) und Umorganisationsempfehlungen getroffen werden, sollte ein ähnlicher Effekt auch an einer noch breiteren Datenbasis gesichert werden können, in die beispielsweise auch psychiatrische Abteilungen mit integriert wären. Da es sich ja um einen Level-2 Effekt, und nicht um ein Patientenmerkmal als Einflussfaktor handelt, könnte der Effekt beim Vergleich mit anderen Organisationsformen noch bedeutsam anders geschätzt werden.

Der letzte zu berichtende Level-2-Effekt ohne Wechselwirkung mit Patientenmerkmalen ist in Tabelle 4.4 als Intercept-Effekt (oder „practice style“) benannt worden. Hier handelt es sich um pauschal bei allen StationspatientInnen beobachtbare, nicht weiter durch die gemessenen Variablen aufklärbare Zu- bzw. Abschläge auf die Verweildauer, wie sie auf der jeweiligen Station als Behandlungsgewohnheit zu be-

obachten waren. Der Mittelwert dieser Stationszuschläge ist natürlich Null. Die Verteilung über alle 135 Stationen ergibt sich aus der nachfolgenden Abbildung.



Insgesamt 24 von 135 verglichenen Stationen erwiesen sich nach Berücksichtigung aller oben erwähnter Patientenmerkmale, und nach Berücksichtigung aller in den Abschnitten 4.2.3 und 4.2.4 berichteten Organisationseffekte dennoch einheitlich bei allen ihren PatientInnen als „Turbo-Behandler“ mit einer teilweise extrem verkürzten Durchschnittsbehandlungsdauer (Intercept-Effekte von bis zu  $-1.79$  log Tagen Verkürzung im Vergleich zum Durchschnitt aller Stationen). 14 Stationen erwiesen sich als „Lang-Brüter“ in dem Sinne, dass auch bei Berücksichtigung aller relevanten Patientenmerkmale der BADO der statistische Erwartungswert der Verweildauer in diesen Stationen bedeutsam höher als der Stationsdurchschnitt war. Wie im Abschnitt 2 erläutert, bedeutet ein solcher Intercept eine einheitliche Verlängerung oder Verkürzung des Stationsaufenthalts für alle dort behandelten PatientInnen. Damit ist er als „Behandlungsgewohnheit“ im eigentlichen Sinne des Wortes interpretierbar.

#### 4.2.4 Stationseffekte in Wechselwirkung mit Patienten-Merkmalen („cross-level interaction“)

Einen letzten Typ von Verweildauer-beeinflussenden Faktoren bilden die sogenannten „cross-level-interactions“. Hierbei wurden zwei organisatorische Merkmale der Stationsspezialisierung als Verweildauer-relevant nachgewiesen: der Anteil an schizophrene PatientInnen, sowie der Anteil an Medikations-frei Behandelten. Während aber der Spezialisierungseffekt bei PatientInnen mit neurotischen Störungen für alle behandelten PatientInnen in der jeweiligen Station gleich wirkte (vgl. obiger Abschnitt), lässt sich der Spezialisierungseffekt bei Spezialstationen für schizophrene

PatientInnen, bzw. bei Stationen mit häufigem Medikations-freien Behandlungsansatz nur für eine Subgruppe von dort behandelten PatientInnen feststellen.

Für PatientInnen mit schizophrenen Störungen bedeutet einerseits die Tatsache ihrer Diagnose einen Verweildauer-verlängernden Effekt. Wie stark sich diese Verlängerung auswirkt, hängt ab von der jeweiligen Station, auf der sie behandelt werden. Diejenigen Stationen, die besonders viele schizophrene PatientInnen im Jahresverlauf behandeln, hospitalisieren dabei genau diese Untergruppe von PatientInnen besonders lange. Dies ist die sprachliche Interpretation des Koeffizienten von 0.45 für die „cross-level-interaction“ aus Diagnose-Schizophrenie und Stationsanteil-Schizophrener. Wird ein nicht-schizophrener Patient auf einer solchen Station behandelt, dann lässt sich für diesen Patienten ein ähnlicher Stationszuschlag bei diesen Stationen durchaus nicht feststellen. Umgekehrt bedeutet die Behandlung eines schizophrenen Patienten auf einer Station mit ausgeprägt niedrigem Diagnosenanteil an Schizophrenie, dass erwartbar seine Verweildauer im Vergleich zu anderen Schizophrenen eher kürzer ausfallen wird.

Die Interpretation des „cross-level-interaction“-Effects für den Anteil Medikations-freier Behandlungen ist wiederum anders. Grundsätzlich wurde ja schon beschrieben, dass grosso modo für PatientInnen, die ohne Psychopharmaka bleiben, mit einer eher verkürzten Verweildauer gerechnet werden kann, bei allerdings unterschiedlichen Effekten über die Stationen hinweg. Werden nun in einer Station aber besonders hohe Anteile von Medikations-freien Behandlungen durchgeführt, oder anders formuliert: handelt es sich um eine Spezialstation, dann wirkt sich bei denjenigen Patientinnen, die dieses Spezialisierungskriterium auch erfüllen (Behandlung ohne Medikamente) der Kontext der Station genau entgegengesetzt zur Verkürzungserwartung aufgrund des Behandlungsmerkmals aus (Koeffizient = + 1.11). Vermutlich werden die beiden unterschiedlichen Patientengruppen, für die ein Medikations-freier Aufenthalt verstärkt registriert wird, in unterschiedlichen Stationen behandelt: Ein Alkoholiker, der nur zur Ausnüchterung seine akute Intoxikationszeit in der Klinik verbringt, also eine extrem kurze Verweildauer in Anspruch nimmt, wird dies vermutlich häufiger in einer Station tun, wo sonst Patienten sind, die sowohl Psychopharmaka erhalten können als auch nicht. Dagegen werden PatientInnen, die ausschließlich psychotherapeutisch (und damit auch länger!) behandelt werden, sich verstärkt auf Stationen aufhalten, die hauptsächlich mit Medikations-freien Psychotherapie-PatientInnen belegt sind. Durch die Freigabe des Schätzparameters „Stationskontext“ konnte diese differierende Bedeutung des Prozessmerkmals „Medikations-freie Behandlung“ in Wechselwirkung mit der jeweiligen Stationsspezialisierung erst korrekt quantifiziert werden.

#### ***4.2.5 Kurzzusammenfassung und Tabellierung der Parameter***

Eine Übersicht über alle Parameter, Schätzfehler und t-Werte (für die statistische Bedeutsamkeit des jeweiligen Parameters) des finalen Modells der logarithmierten Verweildauer (LOS) bei 27.792 Behandlungsepisoden findet sich in Tabelle 4.5. Jeder Effekt auf die Verweildauer wurde danach einsortiert, ob er als Patientenmerkmal entstanden ist (Level-1), oder ob er als Stationsmerkmal zu interpretieren ist (Level 2). In der Spalte „Random Effect“ ist vermerkt, ob die Patientenmerkmale mit Einfluss auf die Verweildauer diesen einheitlich über die Stationen, oder variierend nach Station ausüben. Das Intercept entspricht hier einer globalen Basiserwartung, die aller-

dings in einen globalen Erwartungseffekt, einen klinik-spezifischen Anteil (nicht signifikant, wie oben in 4.2.1 beschrieben), einen stationsspezifischen Anteil (signifikante Varianz, Intercept-Effekt aus Abschnitt 4.2.3), sowie einen personenspezifischen Anteil (üblicher statistischer Messfehler  $\epsilon$ ) aufgespalten wurde.

Unter den Patientenmerkmalen sind aus theoretischen Gründen zwei Variablen im Modell verblieben, für die die übrigen Parameter adjustiert wurden, auch wenn kein nachweislicher Effekt dieser Patientenmerkmale bestand: Das Geschlecht der PatientInnen wurde als „Standard-Konfounder“ in der Epidemiologie in das Modell aufgenommen. Der Effekt der Eingangs-GAF der PatientInnen zeigte sich in früheren Analysen (Frick et al., 1999) als bedeutsam. Im vorliegenden Datensatz wirkte sich die Eingangs-GAF nicht unmittelbar als Level-1 Effekt auf die Verweildauer aus, wurde aber im Modell belassen, damit die Interpretation des „Netto-Effekts“ beim Zusammenhang von Verweildauer und GAF-Zugewinn erhalten blieb: Es sollte die Wirkung eines in der stationären Behandlung „erworbenen“, zusätzlichen sozialen Funktionspunktes in der GAF unabhängig vom Ausgangsniveau am Beginn der stationären Behandlung evaluiert werden.

Die letzten drei Spalten von Tabelle 4.5 („Final explanatory model for LOS“) weisen diejenigen statistischen Tests aus, an denen entschieden wurde, einen Level-1 Effekt als Random-Effekt zu formulieren. T-Werte vom Betrag her größer als der Wert 2 wurden als Entscheidungsrichtlinie für diesen Schritt der Modellierung und bei allen übrigen Entscheidungen über Stationseffekte ( $n=135$ ) benutzt. Bei Entscheidungen über Level-1 Effekte war das Kriterium wegen der ungleich höheren Fallzahl deutlich restriktiver mit  $|t| > 6$  angesetzt (vgl. oben, Abschnitt 2), um einen Mindestbeitrag von 1,5% gemeinsamer Varianz zwischen Prädiktor und Kriterium zu gewährleisten.

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

<b>Table 4.5 Final explanatory model for LOS</b>										
Level-1 predictor variable	Random effect?	Level-2 predictor variable	Coefficient	Standard Error	T value	Variation level (coefficient)	Variance of $u_{jk}, e_{ijk}$	Standard Error	T variance component	
Intercept	yes	Global intercept	2.53136	.10294	24.591	<b>Global Effect <math>\beta_0</math></b>				
							<b>Hospitals <math>v_{0k}</math></b>	0.015310	0.01148	1.334
							<b>Wards <math>u_{0jk}</math></b>	0.453100	0.07760	5.839
							<b>Patients <math>e_{0ijk}</math></b>	1.260130	0.01079	116.787
		Number of annual patients	-0.00023	.00011	-2.091	<b>Wards</b>				
		Proportion of patients with neurotic disorders	0.50965	.19627	2.597	<b>Wards</b>				
Age at admission	yes	Intercept level 2 for slope	0.00357	.00102	3.500	<b>Wards <math>u_{1jk}</math></b>	0.000070	0.00002	3.500	
Readmission? (yes/no)	no		-0.20485	.02012	-10.181	<b>Patient</b>				
Sex (male = 1, female = 0)	no		-0.01197	.01583	-0.756	<b>Patient</b>				
GAF-score at admission	no		0.00082	.00067	1.224	<b>Patient</b>				
Adverse event under pharmacotherapy	no		0.45470	.04143	10.975	<b>Patient</b>				
Complementary services arranged for aftercare	no		0.53549	.02028	26.405	<b>Patient</b>				
Outpatient aftercare (yes/no)	no		0.13838	.01712	8.083	<b>Patient</b>				
Diagnosis: affective disorder	no		0.44090	.02926	15.068	<b>Patient</b>				
Diagnosis: schizophrenic disorder	yes	Intercept level 2 for slope	0.19423	.06200	3.133	<b>Wards <math>u_{12jk}</math></b>	0.100540	0.02202	4.566	
Interaction of diagnosis = schizophrenia and respective proportion within ward	yes	Slope effect of proportion of schizophrenic patients	0.45450	.13158	3.454	<b>Wards</b>				
No drug treatment while hospitalized	yes	Intercept level 2 for slope	-0.95882	.08838	-10.849	<b>Wards <math>u_{8jk}</math></b>	0.193610	0.03419	5.663	
Interaction of no drug treatment and respective proportion within ward	yes	Slope effect of proportion of patients without psychotropic drug treatment	1.11460	.21644	5.150	<b>Wards</b>				
Improvement of GAF score	no		0.01092	.00056	19.500	<b>Patient</b>				
Number of missing values	yes	Intercept level 2 for slope	0.06722	.00838	8.021	<b>Wards <math>u_{10jk}</math></b>	0.005970	0.00105	5.686	

### 4.3 Diskussion der Stationseinflüsse auf die Verweildauer

Krankenhäuser können als Pflichtversorgungskrankenhäuser die Zahl und Art der zu behandelnden PatientInnen nur in einem geringen Ausmaß beeinflussen: über geplante Wiederaufnahmen, und –falls sie über eine eigene Ambulanz verfügen– über Einweisungen aus der eigenen Ambulanz. Den größten Anteil ihres Patientenaufkommens nehmen sie allerdings als „spontane Populationsmorbidity“ entgegen, die sie zu versorgen haben. Zum Vergleich von verschiedenen Institutionen muss daher derjenige Anteil, den die Patienten an der jeweiligen Zielgröße verursachen, statistisch mit geeigneten Mitteln adjustiert werden.

Beim Vergleich der 8 psychiatrischen Kliniken betrug der von den PatientInnen an der Verweildauer verursachte Varianzanteil fast 82%. Die Kliniken selbst banden keinerlei systematischen Varianzanteile. Aber immerhin fast 18% der Verweildauervarianz war der Ebene der Stationen zuzurechnen. Wir haben daraus die Forderung abgeleitet, bei der Qualitätssicherungsdiskussion „den analytischen Blick“ zu refokussieren, und von der Ebene der Krankenhäuser eine Ebene tiefer auf die Beschaffenheit und Leistungsfähigkeit der einzelnen Stationen verstärkt zu achten.

Zu diesem Zweck haben wir die Verantwortlichen aus den beteiligten acht Kliniken zu einem Workshop eingeladen (am 9. Oktober 2001), um in gemeinsamer Diskussion die Bedeutung der neuartigen Erkenntnisse über die Rolle der jeweiligen Stationen bei der Bestimmung der Verweildauer auszuloten. Aus 7 der 8 Kliniken waren Vertreter zu diesem Workshop anwesend.

In Ergänzung der im Abschnitt 4.2 dargestellten stationsspezifischen Effekte erhielten alle Teilnehmer zur Vorbereitung eine Darstellung der oben dargestellten „caterpillar plots“ mit einer pro Klinikum getrennt ausgewiesenen Benennung der konkreten Stationen. Dabei wurden jeweils nur die statistisch auffälligen Stationen dargestellt.

Am Beispiel der (hier anonymisierten) Darstellung zweier Stationen aus dem Klinikum C sei erläutert, welcher Zusammenhang zwischen den stationsspezifischen Auffälligkeiten gefunden wurde: Station 302 (blaues Rautensymbol) behandelt das Aufnahmealter der PatientInnen bei den Entlassungsentscheidungen als eine signifikant geringer wirkend Verweildauer-verlängernde Variable im Vergleich mit den übrigen Stationen. Sie erreicht den insgesamt 6.niedrigsten Wert für den stationsspezifischen Abweichungskoeffizienten  $u_{ijk}$  beim Aufnahmealter. Auch bei den stationsspezifischen Abweichungen bezüglich der Stations-spezifischen Abweichungen für die Diagnosengruppe Schizophrenie erreicht die Station 302 einen extrem niedrigen Wert (3.niedrigster Wert aus 135 Stationen). Genau umgekehrt verhält sich Station 302 bei der Variable „Medikations-freie“ Behandlung: Hier gehört die Station zu denjenigen, die eine „Medikations-freie Behandlung“, die sich grundsätzlich Verweildauer-verkürzend auswirkt, stationsspezifisch doch bei hohem Anteil von Patienten ohne Psychopharmaka wiederum eher Verweildauer-verlängernd behandeln. Auch das Grundniveau der Verweildauer (Intercept-Abweichung) ist in Station 302 statistisch auffällig höher als im Schnitt der Stationen (vgl. die Abbildungen 4.3.3 und 4.3.4).

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Abb. 4.3.1 Spezifische Abweichung Station 302 beim Aufnahmealter

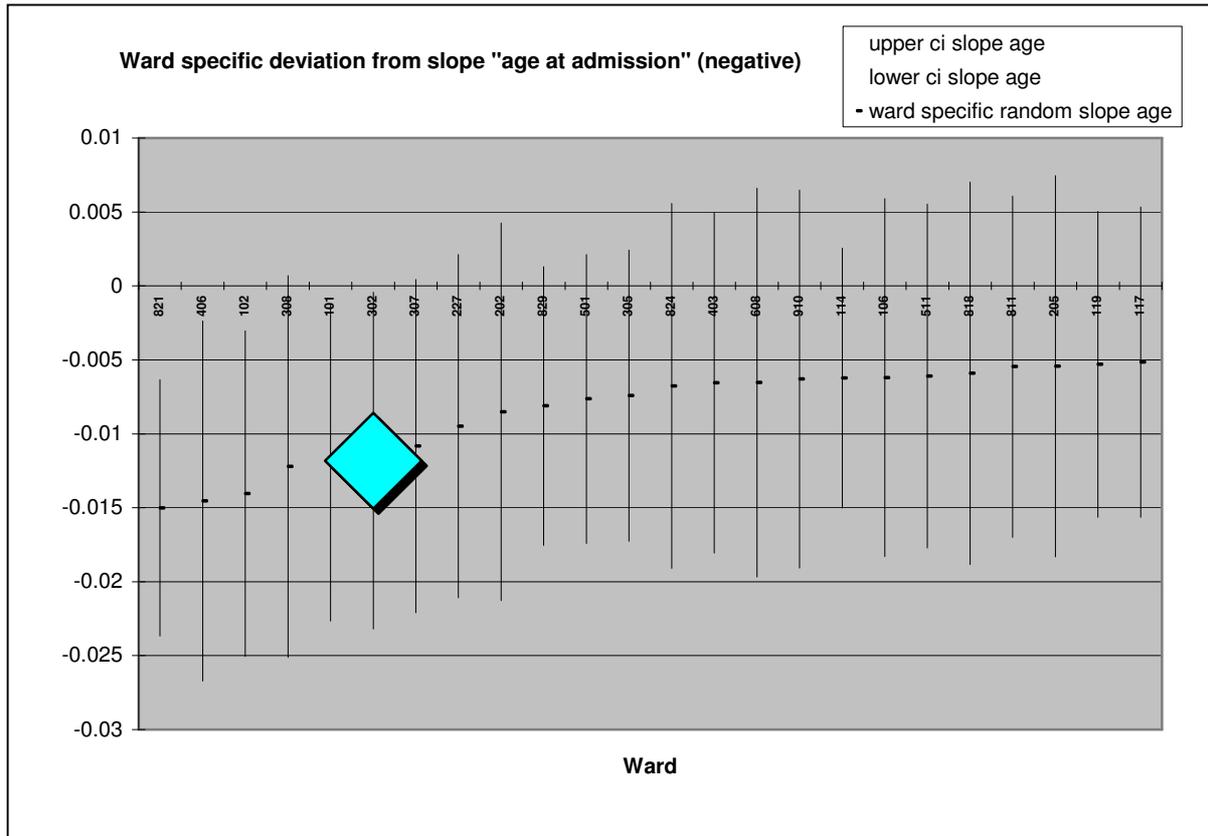


Abbildung 4.3.2 Spezifische Abweichung Station 302 bei Diagnose Schizophrenie

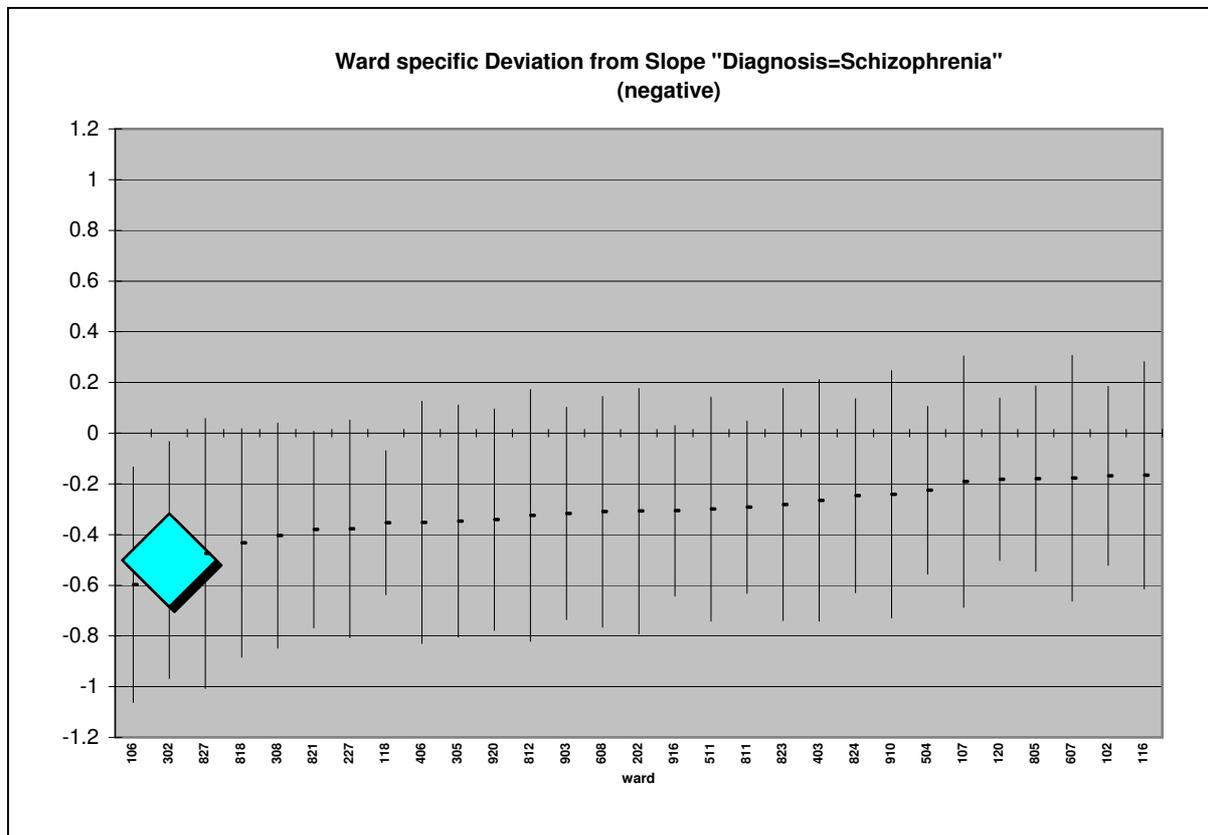


Abb. 4.3.3 Spezifische Abweichung Station 302 bei „Medikationsfreie Behandlung“

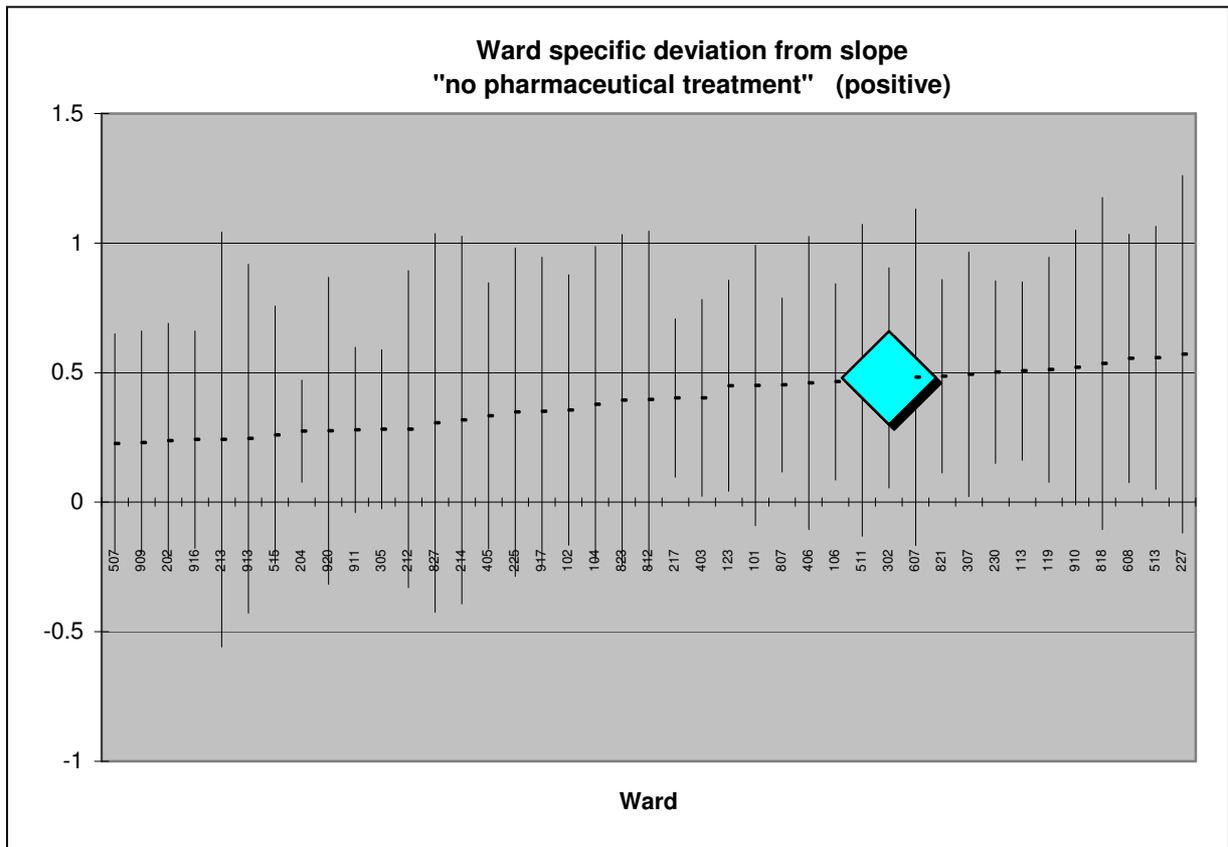
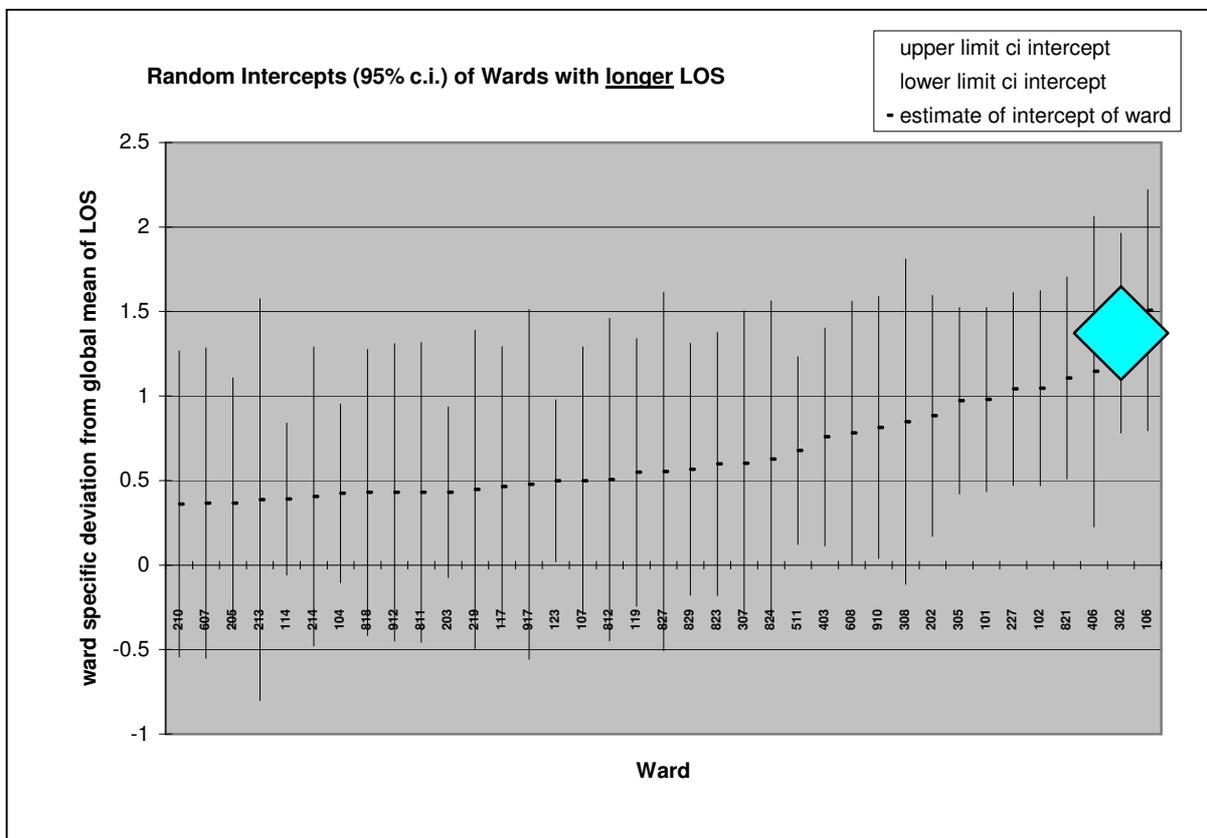


Abb.4.3.4 Spezifische Abweichung Station 302 bei Intercept



Station 302 veranschaulicht in ihrem Verhalten sehr typisch die systematischen Zusammenhänge, die insgesamt über alle Stationen hinweg bei den stationsbedingten Varianzanteilen der Verweildauer zu beobachten sind:

**Zusammenhang der Stationsbesonderheiten bei n=135 Stationen**

Stationsspezifische Abweichung:		U ANZMV	U SCHIZO	u aufalter	u intercept	u nopharm
		Anzahl der missing values	Diagnose: Schizophrenie	Alter bei Aufnahme	"practice style"	Medikationsfreie Behandlung
U_ANZMV	Korrelation nach Pearson	1.000	-0.198	0.117	0.015	-0.001
	Signifikanz (2-seitig)	,	0.021	0.176	0.860	0.989
	N	135	135	135	135	135
U_SCHIZO	Korrelation nach Pearson	-0.198	1.000	0.598	-0.806	-0.543
	Signifikanz (2-seitig)	0.021	,	0.000	0.000	0.000
	N	135	135	135	135	135
U_aufalter	Korrelation nach Pearson	0.117	<b>0.598</b>	1.000	-0.879	-0.509
	Signifikanz (2-seitig)	0.176	0.000	,	0.000	0.000
	N	135	135	135	135	135
U_intercept	Korrelation nach Pearson	0.015	<b>-0.806</b>	<b>-0.879</b>	1.000	0.640
	Signifikanz (2-seitig)	0.860	0.000	0.000	,	0.000
	N	135	135	135	135	135
U_nopharm	Korrelation nach Pearson	-0.001	<b>-0.543</b>	<b>-0.509</b>	<b>0.640</b>	1.000
	Signifikanz (2-seitig)	0.989	0.000	0.000	0.000	,
	N	135	135	135	135	135

Um der Frage nach einer möglichen Systematik bei den Abweichungen der Stationen nachzugehen, wurden auf dem Stationslevel die geschätzten  $u$ -Terme pro Station (siehe Gleichungen (3) und (4) im Abschnitt 2.1) miteinander korreliert. Das Ergebnis (vgl. Tabelle oben) weist darauf hin, dass die konkrete Station 302 durchaus keinen Einzelfall darstellt, sondern dass die Konstellation „generell lang–verkürzt bei Schizophrenie–lang bei medikationsfreien Patienten–das Alter weniger gewichtend“ ein gängiges Muster der Abweichung darstellt: Stationen, die (wie Station 302) generell länger behandeln ( $u_{intercept}$ ), tun dies auch verstärkt ( $r = 0.64$ ) bei einem höheren Anteil von Medikations-freien PatientInnen für genau diese PatientInnen ( $u_{nopharm}$ ). Dagegen ist erwartbar, dass sie ( $r = - 0.51$ ) das Aufnahmealter weitaus weniger Verweildauer-verlängernd berücksichtigen als der Stationsdurchschnitt ( $u_{aufalter}$ ), und auch die Diagnose Schizophrenie ( $r = - 0.54$ ) weitaus weniger deutlich als Verlängerungsfaktor berücksichtigen ( $u_{schizo}$ ).

Den Gegenpol zu diesem Muster bildet ein Verhalten, wie es im selben Krankenhaus die Station 304 an den Tag gelegt hat. Diese Station behandelt generell eher kürzer als der Stationsdurchschnitt (Abb. 4.3.5, rotes Dreieck als Symbol, zweitniedrigster Wert insgesamt) und berücksichtigt auch die Medikations-freien Patienten noch stärker als Verweildauer-verkürzend als andere Stationen (Abb. 4.3.6). Dagegen wirkt sich in Station 304 die Diagnose „Schizophrenie“ überdurchschnittlich Verweildauer-steigernd aus (Abb.4.3.7, dritthöchster Wert insgesamt), ebenso wie das Lebensalter der PatientInnen (Abb.4.3.8 höchster „Zuschlag“ unter allen Stationen).

Abb. 4.3.5 Spezifische Abweichung Station 304 beim Intercept

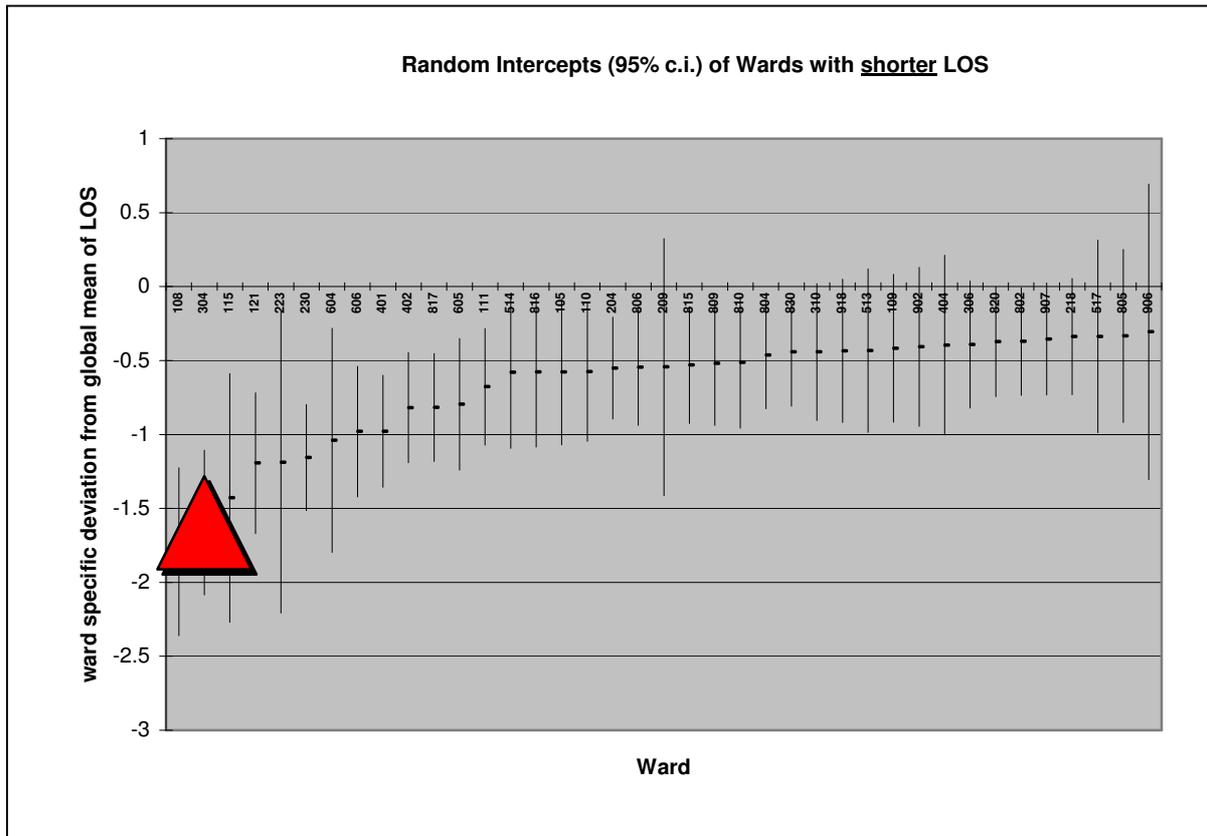


Abb. 4.3.6 Spezifische Abweichung Station 304 „Medikations-freie Behandlung“

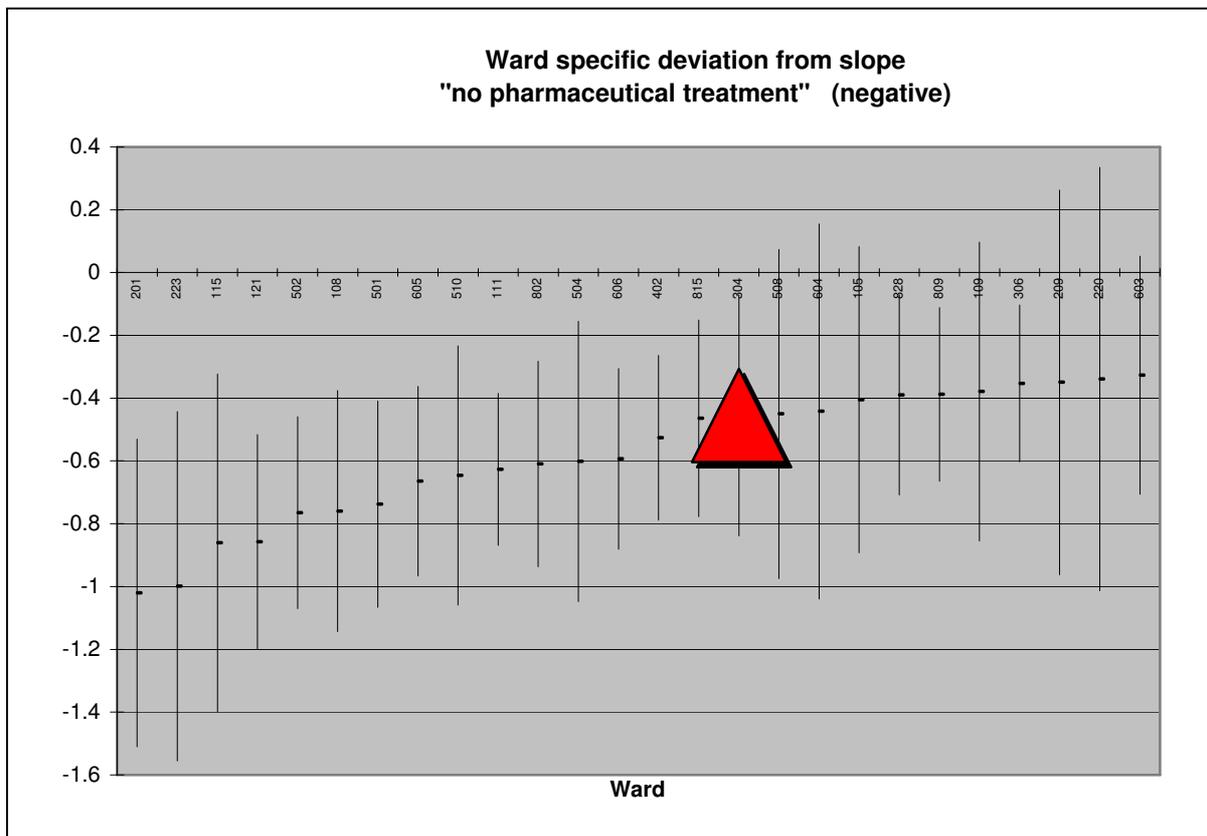


Abb. 4.3.7 Spezifische Abweichung Station 304 „Diagnose Schizophrenie“

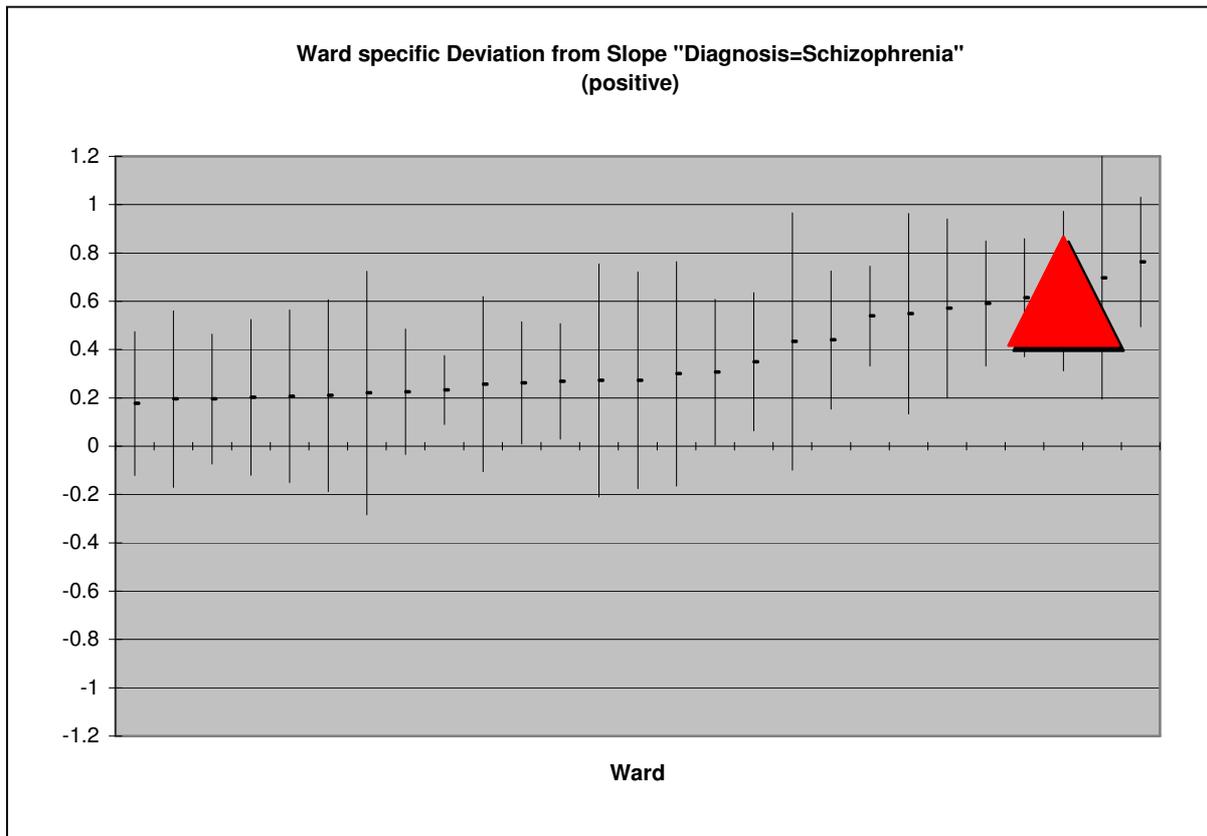
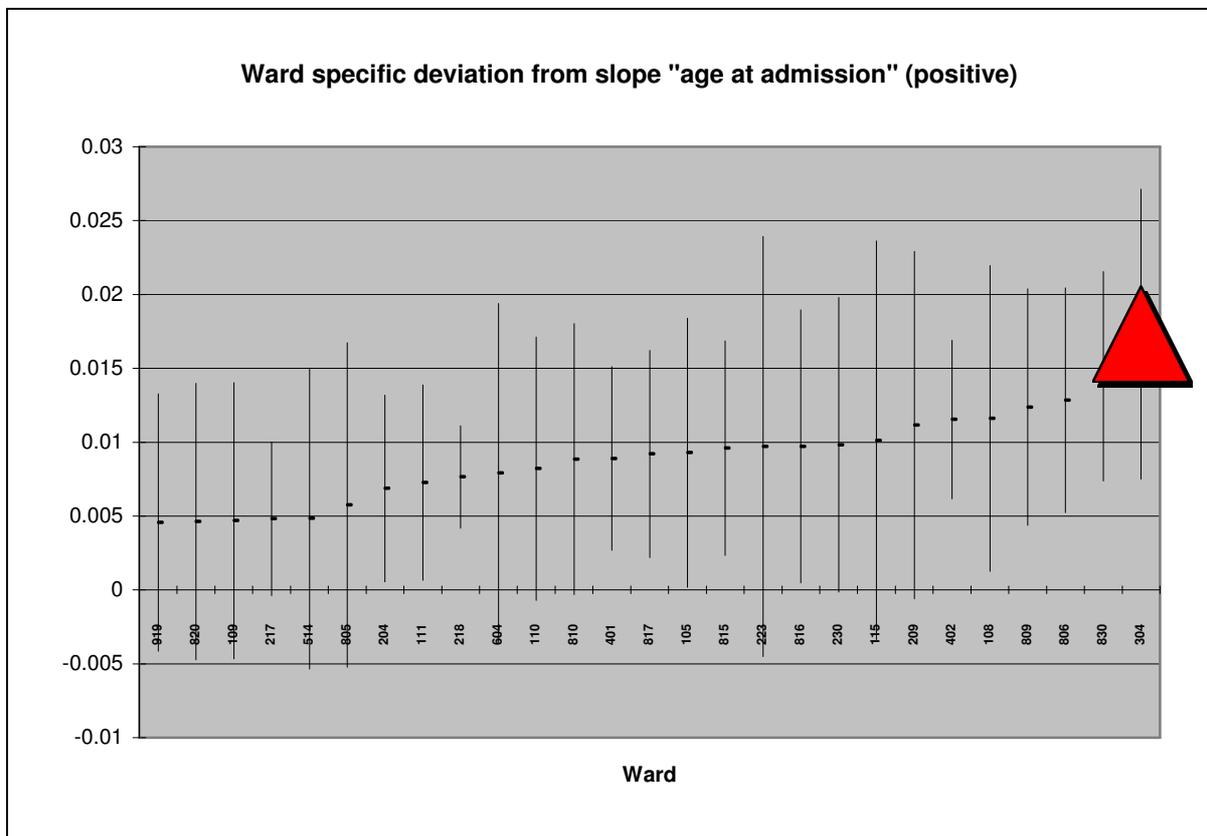


Abb. 4.3.8 Spezifische Abweichung Station 304 „Aufnahmealter“



Der fünfte im finalen statistischen Modell für die Verweildauer vorgefundene stationspezifische Effekt, der random-slope-Effekt für den Zusammenhang der Dokumentationsdichte mit der Verweildauer, zeigte bei der Auswertung der Kennwerte für die 135 Stationen keinerlei bedeutsamen Zusammenhang ( $|r| > 0.3$ ) mit den übrigen Stationsabweichungen. Dies deutet darauf hin, dass der Zweck der Kontrolle eines potenziell informativen Zensierungsmusters erfüllt ist, und sich in der Dokumentationsdichte dann keine spezifischen Organisationsprinzipien von Stationen mehr verbergen, die kovariieren würden mit den übrigen 4 Variablen.

In der Diskussion der stationsbezogenen Ergebnisse beim Workshop am 9. Oktober 2001 fiel den Teilnehmern auf, dass sehr viele der auffälligen Stationen vom Muster der Station 304 (rote Dreiecke) Aufnahmestationen sind, die hier (eher atypisch) auch als Entlassungsstationen fungieren. Die auch im Vergleich zu anderen Stationen extrem kurzen Verweildauern bei medikationsfreien Behandlungen wie auch der generell sehr niedrige Verweildauer-Sockel (Intercept-Effekt) dieser Stationen könnten nach Meinung der Workshop-Teilnehmer interpretiert werden als „nahezu-Fehlallokationen“<sup>3</sup> von PatientInnen in stationäre Behandlung: Oft werden hier PatientInnen nur 1 oder 2 Tage zu einer Beobachtung (oder Ausnüchterung) stationär aufgenommen; eine weitere stationäre Behandlung erweist sich dann aber nicht als notwendig, und die PatientInnen werden aus der Aufnahmestation wieder entlassen.

Demgegenüber erklärten die Diskussionsteilnehmer die überproportionalen Verweildauern bei älteren und schizophren erkrankten PatientInnen bei der Entlassung von Aufnahmestationen mit einem „Entmischungseffekt“ in dem Sinne, dass es sich hier wohl um extrem schwer Kranke handeln könnte, die für eine sonst übliche, Klinik-interne Verlegung auf Therapiestationen mit Entlassungsorientierung nicht in Frage gekommen seien wegen ihres Krankheitsbildes.

Zur Deutung des Typs nach dem Muster von Station 302 (blaue Raute als Symbol) gewannen die Vertreter der Kliniken den Eindruck, dass hier verstärkt Psychotherapiestationen identifiziert wurden, die länger als der Gesamtdurchschnitt behandelten, die Medikations-freie Behandlung daher eher Verweildauer verlängernd betrachteten, aber den sonst üblichen „Alterszuschlag“ oder „Schizophrenie-Malus“ weniger deutlich handhaben würden, wenn denn überhaupt ein älterer oder schizophrener Patient auf einer solchen Station behandelt und von dort entlassen werde. Möglicherweise würden schizophrene PatientInnen dort auch kaum behandelt, sondern schnell „weitergereicht“, wenn sonst das Stationskonzept gefährdet wäre, so der Eindruck.

Aus dem internen Wissen um die Organisationen äußerten die Klinikvertreter bei manchen der als auffällig identifizierten Stationen auch die Meinung, es seien durch die Multi-Level-Analysen persönliche Entscheidungsstile des/der einen oder anderen Oberarztes / -ärztin herausgearbeitet worden, die nicht weiter durch Organisationsbesonderheiten, sondern eher durch einen besonders „fürsorglichen“ oder „robusten“ Entscheidungsvorgang gekennzeichnet wären.

Dass in Tabelle 4.5.1 der Methodenfaktor „Anzahl der missing values“, der als stationspezifische Besonderheit der jeweils geübten Dokumentationsdichte eingeführt worden war, keine Level-2-Korrelation mit den übrigen Stationseinflüssen erreicht, kann als Zeichen dafür interpretiert werden, dass – wenn einmal dafür adjustiert

---

<sup>3</sup> Vgl. Frick (2002) zur Diskussion des Begriffs „Fehlallokation“ speziell in der Psychiatrie.

wurde, dass manche Stationen bei bestimmten Verweildauerlängen eher mehr oder eher weniger genau die BADO-Variablen ausfüllen, - keine weiteren Zusammenhänge mit anderen Organisationsspezifika mehr zu finden sind. Der Sinn der Methodenvariable, für stationsspezifische Dokumentationsgewohnheiten zu adjustieren und so die inhaltliche Interpretation der übrigen Variablen zu verbessern, scheint also eingelöst zu sein.

#### **4.4 Methodisches Fazit**

Der Einsatz von Multi-Level-Analysen erweitert die Interpretationsmöglichkeiten entscheidend: Das „Vergleichen“ von Krankenhäusern mittels statistischer Verfahren erleichtern zu wollen, die lediglich die Klassifikation der Krankenhäuser systematischer und fehlerfreier als ein menschlicher Urteiler durchführen können, läßt den zweiten Schritt eines Vergleiches völlig offen, nämlich die Urteilsbildung darüber, wie die dann gebildeten Gruppen von Krankenhäusern hinsichtlich irgendwelcher –noch zu bestimmender – Zielvariablen bewertet werden sollen. Kein Klassifikationsverfahren kann diesen Schritt leisten, wiewohl dieser Ansatz in der medizinischen Fachliteratur sehr beliebt ist (z.B. (Betzler & Haun 1998) (Gerste 1996) (Sahlmüller 2001; Schneeweiss & Sangha 2000)).

In den Fällen, wo beispielsweise Qualitätsindikatoren über Krankenhäuser hinweg verglichen werden sollen, kann man den Vergleich von einem Klassifikationsproblem in ein Regressionsproblem umformulieren und erreicht dadurch eine Aussage, welche Merkmale der Krankenhäuser den jeweiligen Qualitätsindikator beeinflussen. Dass Patienten in dieser Sichtweise ebenfalls zu den „Merkmalen“ eines Krankenhauses gerechnet werden müssen, erforderte die Einführung eines Mehrebenenansatzes bei der Analyse. Damit wird die Perspektive des Vergleiches auf den eigentlichen Adressaten von medizinischen Leistungen refokussiert. Das Krankenhaus bleibt eine – unter vielen anderen – potenzielle Einflussgröße auf das Behandlungsergebnis. Indem in den Multilevel-Analysen organisationelle und individuelle Bedingungsfaktoren methodisch adäquat in Relation zueinander gestellt werden können, ergibt sich nicht nur ein Screening-Verfahren für Krankenhäuser mit möglicherweise schlechterer Qualität, sondern ergeben sich vielmehr auch Hinweise darauf, an welchen organisatorischen Bedingungen denn Veränderungen zur Qualitätsverbesserung im Anlassfalle ansetzen können.

Dies ist ein wichtiger Vorteil des vorgestellten Verfahrens. Als Nachteil muss wohl in Kauf genommen werden, dass Multilevel-Modelle einen deutlich höheren Aufwand sowohl bei der statistischen Durchführung, wie auch bei ihrer Darstellung verlangen als weniger komplexe Vergleichsmethoden.

## 5. Vergleiche der Effektivität und Kosten-Effektivität (Autoren: Winfried Barta, Harald Binder)

### 5.1 Einleitung

Bei einem Vergleich der für den jeweiligen Patientenmix adjustierten mittleren Verweildauer von Krankenhäusern wird mit der Verweildauer zwar ein wichtiger Aspekt des Behandlungsgeschehens untersucht. Der eigentliche Zweck eines stationären Patientenaufenthalts, das Erreichen eines gewissen Therapieerfolgs, bleibt dabei jedoch außen vor. Insbesondere beim Vergleich von Kliniken der somatischen Disziplinen wird diesbezüglich argumentiert, es existiere kein geeignetes, auch über verschiedene Diagnosegruppen und Fächer hinweg vergleichbares Maß für den Therapieerfolg. Daher entziehe sich dieser primäre Outcome einer Krankenhausbehandlung quantitativen Analysen und man müsse sich deshalb beim Klinikvergleich auf sekundäre Outcomes beschränken (vgl. Reichsthaler, 2001, S. 10 sowie S. 50 ff.). Unter diesen spielen insbesondere die Belegtage eines Krankenhauses, also die kumulierten Verweildauern aller behandelten Patienten innerhalb eines gewissen Zeitraumes eine zentrale Rolle.

Innerhalb der Psychiatrie steht jedoch mit dem GAF-Score (global assessment of functioning, Wittchen et al., 1991), einer Einschätzung des sozialen Funktionsniveaus auf einer Ratingskala von null bis hundert, ein für alle psychiatrischen Patienten gleichermaßen anwendbares Outcome-Maß zur Verfügung. Gemäß DGPPN-BADO wird der GAF-Score bei Aufnahme wie bei Entlassung routinemäßig bei allen stationär-psychiatrischen Patienten erhoben. Auch im hier vorliegenden Datensatz der acht psychiatrischen Kliniken steht dieser zur Verfügung (vgl. Kap.3.2.2). Das Ausmaß der individuellen Veränderung (i.d.R. Verbesserung) im GAF-Score zwischen Aufnahme und Entlassung, also  $\Delta\text{GAF} = \text{GAF}_{\text{Entlassung}} - \text{GAF}_{\text{Aufnahme}}$ , kann über alle psychiatrischen Diagnosegruppen hinweg als individueller Therapieerfolg des jeweiligen stationären Aufenthalts betrachtet werden. Es bietet sich daher an, die Krankenhäuser nicht nur bezüglich ihrer (Casemix-adjustierten) mittleren Verweildauer (vgl. Kap.4), sondern auch bezüglich ihres – wiederum geeignet adjustierten – mittleren Therapieerfolgs zu vergleichen. Betrachtet man einen möglichst hohen Therapieerfolg als eigentliches Ziel der stationären Behandlung, so wird damit analysiert, inwieweit die einzelnen Häuser dieses Behandlungsziel erreichen (Effektivität). Setzt man zusätzlich den erreichten Therapieerfolg in Bezug zu den dafür aufgewandten Kosten, so ergeben sich vergleichende Aussagen zur Kosten-Effektivität der Behandlungen in den entsprechenden Krankenhäusern. Da in der Krankenhausfinanzierung in der Psychiatrie tagesgleiche Pflegesätze nach wie vor eine wichtige Rolle spielen, wird als einfacher Kostenindikator die Verweildauer betrachtet. Vorläufer der hier vorgestellten Analysen wurden bereits in Barta et al. (2003) berichtet.

### 5.2 Fragestellung

1. Für den Effektivitäts-Vergleich soll untersucht werden, ob sich die Krankenhäuser - nach Adjustierung für Patientenvariablen – im durchschnittlich erzielten Therapieerfolg (Delta-GAF) unterscheiden. Da die acht Krankenhäuser hier nicht als Zufalls-

stichprobe aus einer zugrundeliegenden Krankenhaus-Population betrachtet werden, und da hier nicht eine solche zugrundeliegende Population interessiert, sondern vielmehr Aussagen über genau die betrachteten acht Häuser gemacht werden sollen, wird lediglich ein Zwei-Ebenen-Modell (Patienten in Stationen) verwendet, wobei die Krankenhäuser als feste Effekte (modelliert mit Dummy-Variablen) auf Ebene zwei berücksichtigt werden. Auf diese Weise können auch tatsächliche Unterschiede zwischen den Häusern leichter gefunden werden, da kein „shrinkage“ – eine Verkleinerung der Krankenhauseffekte durch Annäherung an ihren Mittelwert – der entsprechenden Effekte stattfindet (vgl. Goldstein, 1995).

Die Casemix-Adjustierung erfolgt dabei für alle Stationen bzw. Krankenhäuser auf gleiche Weise (feste Effekte), es werden (auch für die Verweildauer) keine unterschiedlichen Steigungen (random slopes) zugelassen. Der Grund dafür ist, dass sonst ein Overall-Vergleich der Krankenhäuser mit *einer* Effektivitäts-Kennzahl für jedes Haus deutlich erschwert würde. Denn um zu verhindern, dass tatsächliche Krankenhausunterschiede in der Verteilung der Stationseffekte „untergehen“ (dass etwa bei einem random slope für Alter das Gros der Stationen eines Hauses am linken Rand der Verteilung dieser Stationseffekte liegt), also nicht mehr gefunden werden können, müsste der entsprechende Random-Slope-Effekt zusätzlich krankenhausspezifisch modelliert werden (über Variablen KH1 x Alter, KH2 x Alter, etc.). Das aber bedeutet, dass ein Vergleich der Häuser nur mehr altersspezifisch erfolgen kann: Es könnte sich etwa ergeben, dass in der Altersgruppe der 30-40-jährigen Krankenhaus A im Mittel mehr Therapieerfolg erzielt, als Krankenhaus B, dass dies in der Altersgruppe der 50-60-jährigen aber genau umgekehrt ist. Auch wenn solche Aussagen sicherlich von Interesse sein können, soll hier zugunsten eines leichter interpretierbaren Overall-Vergleichs darauf verzichtet werden.

2. Für Aussagen zur Kosten-Effektivität wird dann zusätzlich ein random slope für die (logarithmierte) Verweildauer, also unterschiedliche Verweildauer-Steigungen (Grenznutzen) für die verschiedenen Stationen bzw. Krankenhäuser zugelassen. Die Modellierung der (festen) Effekte der Krankenhäuser erfolgt dabei über zusätzliche Interaktionsterme (KH1 x Verweildauer, KH2 x Verweildauer, etc.). Zum Vergleich der Krankenhäuser muss dann simultan deren Achsenabschnitt und ihre Verweildauer-Steigung (bzw. die jeweilige Abweichung vom Mittel) betrachtet werden. Die Verweildauer wurde hier logarithmiert betrachtet, da auf diese Weise eher ein *linearer* Effekt auf den Therapieerfolg erwartet werden kann: Bei einer Verweildauerverlängerung von 20 auf 21 Tage ist nicht der selbe Zugewinn im Therapieerfolg zu erwarten, wie bei einer Verlängerung von 5 auf 6 Tage. Genau das wird jedoch bei einem linearen Zusammenhang von Verweildauer und Therapieerfolg unterstellt. Ein linearer Zusammenhang zwischen *logarithmierter Verweildauer* und Therapieerfolg dagegen impliziert den selben Zuwachs im Therapieerfolg jeweils bei Verlängerung der *logarithmierten Verweildauer* um eins, was einer Verlängerung der *Verweildauer* um den Faktor 2,7 entspricht. Hier wird also beispielsweise der selbe Zuwachs im Therapieerfolg erwartet beim Übergang von 20 auf 54 Tage, wie beim Übergang von 5 auf 14 Tage. Dies ist für stationär-psychiatrische Klinikaufenthalte eine deutlich realistischere Annahme. Mit dieser zweiten Analyse kann die Frage beantwortet werden, ob neben möglichen Unterschieden im mittleren adjustierten Therapieerfolg beispielsweise in Krankenhaus A ein *zusätzlicher* Tag Verweildauer mehr *zusätzlichen* Therapieerfolg erbringt, als in Krankenhaus B.

## Modell-Selektion

Die Modellkonstruktion erfolgt in mehreren Schritten. Aus theoretischen Erwägungen werden die Variablen Alter, Geschlecht, Diagnose, (logarithmierte) Verweildauer, Aufnahme-GAF, sowie Anzahl fehlender Werte (als Indikator für die Dokumentationsqualität) a priori ins Modell aufgenommen. Die Variablen Aufnahmealter, Aufnahme-GAF und logarithmierte Verweildauer werden dabei auf ihren Gesamt-Mittelwert zentriert (grand-mean-centering). Dann erfolgt schrittweiser Variableneinschluss mittels Wald-Test unter den verbleibenden Variablen, wobei zunächst nur lineare Terme ohne Interaktionen betrachtet werden. Mögliche nicht-lineare Zusammenhänge werden zunächst mit nicht-parametrischen Verfahren (u.a. „generalized additive models, Hastie & Tibshirani, 1990) exploriert. Die aus diesen Untersuchungen stammenden Hinweise auf nicht-lineare Effekte und Interaktionen werden dann mit Hilfe von polynomialen und/oder Interaktions-Termen in das bisherige Modell integriert. Diese Erweiterungen werden anhand des Informationskriteriums von Akaike (AIC) abgesichert, um einerseits „overfitting“ zu verhindern und andererseits wichtige Terme nicht versehentlich zu ignorieren.

## 5.3 Ergebnisse

### Varianzanteile

Zunächst ergibt sich, dass 91.6 % der Varianz des Therapieerfolgs auf Patienten-Ebene, 6.8 % auf Stationsebene und 1.6 % auf Klinikebene zu verorten sind. Ein Devianztest gegen ein Zwei-Ebenen-Modell ergibt  $p=0.00015$ , d.h. diese 1,6 % unterscheiden sich signifikant von Null. In einem ersten Schritt führte schrittweiser Variableneinschluss linearer Regressoren ohne Interaktionen zu folgendem Modell (Abbildung 5.1):

```
Therapieerfolg = Baseline
+ logarithmierte Verweildauer
+ Aufnahmealter
+ Aufnahme-GAF
+ Geschlecht
+ Zahl der fehlenden Werte
+ Diagnosegruppe
+ Erstaufnahme
+ Rehospitalisierung binnen 30 Tagen
+ Bildung
+ Lebt mit (Ehe-)Partner zusammen
+ Voll-/Teilzeit Beschäftigung vor Aufnahme
+ (teil-)stationäre Weiterbehandlung vereinbart
+ komplementäre/ähnliche Weiterversorgung vereinbart
+ ambulante Weiterbetreuung vereinbart
+ Weiterbehandlung durch Allgemeinarzt vereinbart
+ Klinikeffekt
+ zufällige Stationseffekte
+ Fehlerterm
```

Abb. 5.1: Erstes vorläufiges Vorhersagemodell für den Therapieerfolg.

## **Nichtparametrische explorative Analysen**

Mit Hilfe von „generalized additive models“ werden im Folgenden in verschiedenen Modellen die interessierenden metrischen Variablen - Alter, Aufnahme-GAF und (logarithmierte) Verweildauer - auf nicht-lineare Effekte und Interaktionen untersucht. Die hierarchische Datenstruktur (der Stationseffekt) wird hierbei vorerst nicht berücksichtigt, da „generalized additive models“ in ihrer herkömmlichen Formulierung keine zufälligen Effekte zulassen, wie sie für die Modellierung der Stationseffekte benötigt würden. Da es sich hier allerdings nur um eine explorative Analyse handelt, die nur Anregungen für die spätere Konstruktion eines Modells mit zufälligen Stationseffekten geben soll, wird dieses Problem nicht weiter beachtet.

Betrachtet man zunächst ein globales Modell, das für alle drei metrischen Kovariaten glatte Effekte zulässt, allerdings keine Interaktionsmöglichkeiten vorsieht (siehe Abbildung 5.2), so finden sich keine ausgeprägten nichtlinearen Effekte (siehe Abbildung 5.3). Inhaltliche Überlegungen legen allerdings nahe, dass die Effekte für unterschiedliche Diagnosegruppen deutlich unterschiedlich ausfallen dürften. Dies könnte dazu führen, dass sich bei einer Analyse über alle Diagnosegruppen hinweg kein Effekt mehr findet. Deshalb wird die Analyse im Folgenden für einzelne Diagnosegruppen separat durchgeführt. Gegen ein globales Modell mit Diagnose-spezifischen glatten Effekten spricht vor allem die Unhandlichkeit derart großer Modelle.

Therapieerfolg = Baseline  
+ **glatter Effekt der Verweildauer**  
+ **glatter Effekt des Aufnahmealters**  
+ **glatter Effekt des Aufnahme-GAF**  
+ parametrischer Effekt von Geschlecht,  
Zahl der fehlenden Werte, Diagnosegruppe, Erstaufnahme, Rehospitalisierung, Bildung,  
Partnerschafts- und Arbeitsverhältnissen,  
Art der vereinbarten Weiterbetreuung  
+ **Klinikeffekt**  
+ Fehlerterm

Abb. 5.2: Vorhersagemodell für den Therapieerfolg mit nichtparametrischem Einfluss der Verweildauer, des Alters bei Aufnahme und des GAF bei Aufnahme und mit weiteren parametrischen Komponenten.

Nach Diagnose-spezifischer Betrachtung von separaten glatten Effekten der metrischen Kovariaten, die deutlichere (nicht-lineare) Effekte zeigen, werden glatte Interaktionen analysiert. Beispielhaft seien nachfolgend zwei Modelle für die Diagnosegruppe „Schizophrenie“ angeführt:

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

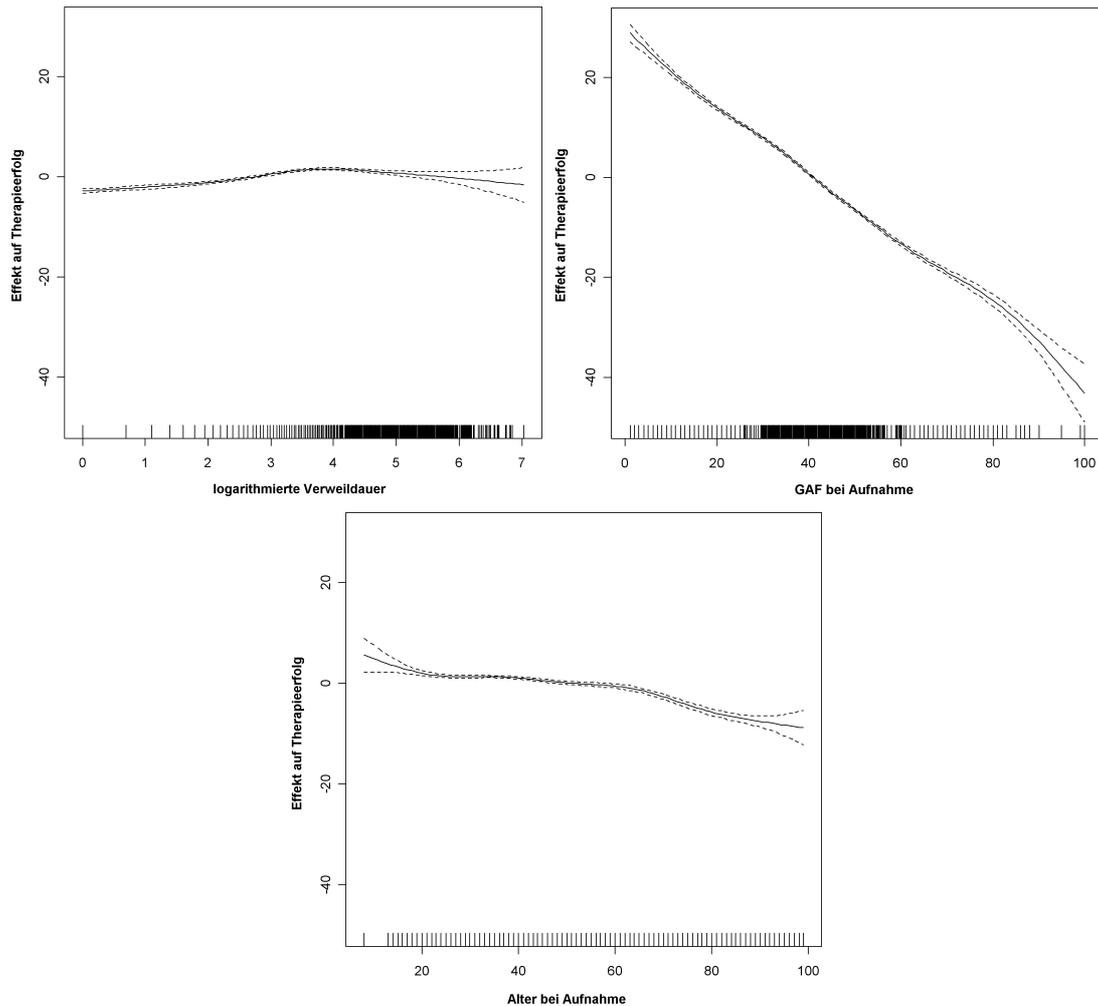


Abb. 5.3: Glatte Effekte für die metrischen Kovariaten in einem globalen Modell ohne Interaktionen.

Therapieerfolg = Baseline  
 + **gemeinsamer glatter Effekt der Verweildauer und des Aufnahmealters**  
 + **glatter Effekt des Aufnahme-GAF**  
 + parametrischer Effekt von Geschlecht,  
 Zahl der fehlenden Werte, Diagnosegruppe,  
 Erstaufnahme, Rehospitalisierung, Bildung,  
 Partnerschafts- und Arbeitsverhältnissen,  
 Art der vereinbarten Weiterbetreuung  
 + **Klinikeffekt**  
 + Fehlerterm

Abb. 5.4: Vorhersagemodell für den Therapieerfolg in der Diagnosegruppe „Schizophrenie“ mit gemeinsamen nicht-parametrischem Einfluss der Verweildauer und des Alters bei Aufnahme und nicht-parametrischem Einfluss des GAF bei Aufnahme.

Das erste modelliert den Effekt des Aufnahme-GAFs nicht-linear (glatt) und lässt einen gemeinsamen, nicht-linearen (glatten) Effekt der Variablen Verweildauer (loga-

rithmiert) und Aufnahmealter zu (siehe Abbildung 5.4). Abbildung 5.5 zeigt die geschätzten glatten Effekte. Hier gibt vor allem die dreidimensionale Grafik zum gemeinsamen Effekt von Verweildauer und Alter Hinweise auf eine mögliche komplexe Interaktion, wie sie in ein späteres Modell mit Zufallseffekten aufgenommen werden sollte.

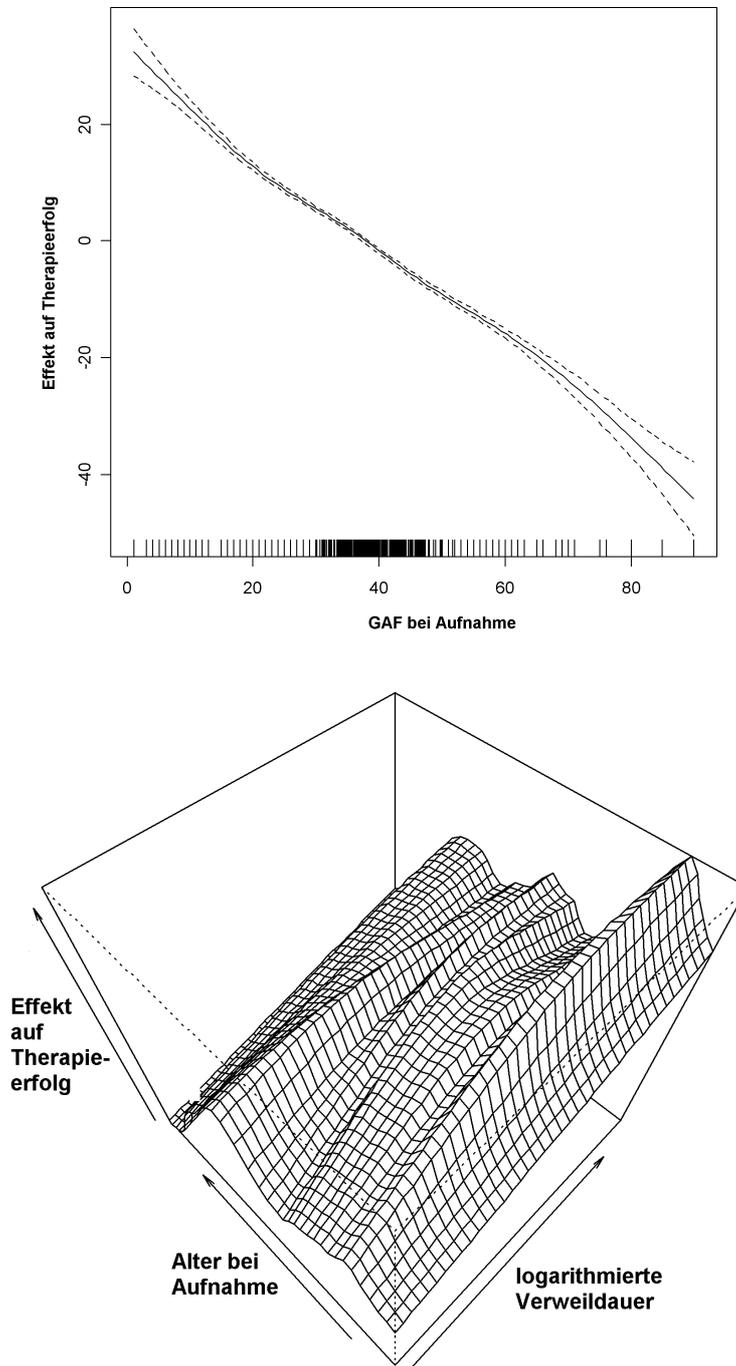


Abb. 5.5: Modell für die Diagnosegruppe „Schizophrenie“ mit nicht-linear modelliertem (glatten) Effekt für die Variable Aufnahme-GAF und gemeinsamem nicht-linearen (glatten) Effekt für Verweildauer und Aufnahmealter.

Das zweite Modell betrachtet den Effekt von Alter und Aufnahme-GAF gemeinsam und den der Verweildauer einfach glatt (siehe Abbildung 5.6). Von den in Abbildung 5.7 gezeigten Schätzungen ist vor allem diejenige für den Effekt der Verweildauer interessant, da sie es nahe legt, die Verweildauer als nicht-lineare Komponente in ein späteres Modell einfließen zu lassen.

Frick, Krischker, Cording (Regensburg)  
**Freiwillige Krankenhausvergleiche zur externen Qualitätssicherung**

Therapieerfolg = Baseline

- + **glatter Effekt der Verweildauer**
- + **gemeinsamer Eff. des Aufnahmealters und des Aufnahme-GAF**
- + parametrischer Effekt von Geschlecht, Zahl der fehlenden Werte, Diagnosegruppe, Erstaufnahme, Rehospitalisierung, Bildung, Partnerschafts- und Arbeitsverhältnissen, Art der vereinbarten Weiterbetreuung
- + **Klinikeffekt**
- + Fehlerterm

Abb. 5.6: Vorhersagemodell für den Therapieerfolg in der Diagnosegruppe „Schizophrenie“ mit nicht-parametrischem Einfluss der Verweildauer und gemeinsamen Einfluss des Alters bei Aufnahme und des GAF bei Aufnahme.

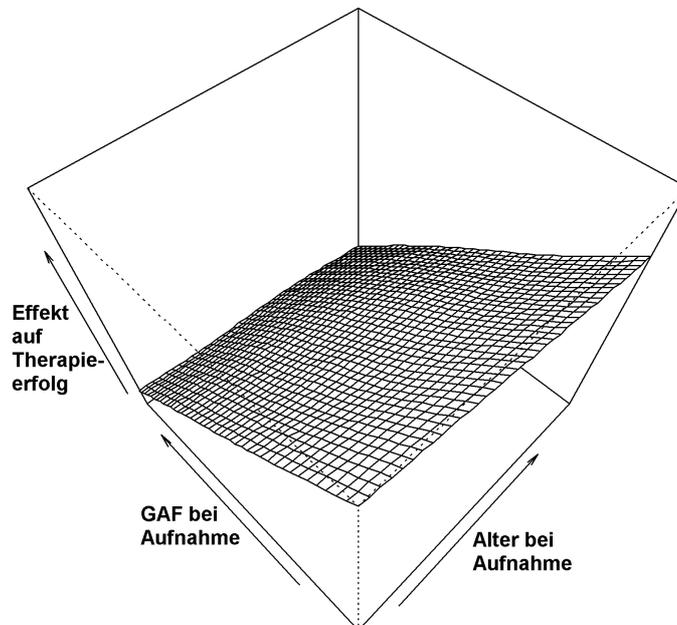
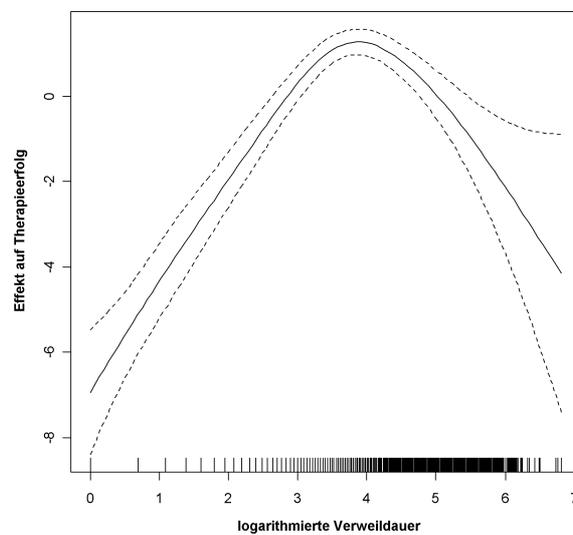


Abb. 5.7: Modell für die Diagnosegruppe „Schizophrenie“ mit glatterm Effekt für die Variable Verweildauer und gemeinsamen glatten Effekt für Aufnahme-GAF und Aufnahmealter

Durch Betrachtung der Ergebnisse für die weiteren Diagnosegruppen lassen sich so weitere Indizien für nicht-lineare Effekte gewinnen.

## Hierarchische Modelle zur Analyse von Effektivität und Kosten-Effektivität

Das im ersten Schritt durch schrittweisen Variableneinschluss konstruierte Modell wird nun aufbauend auf den vorgestellten explorativen Analysen durch nichtlineare Komponenten und Interaktionen erweitert. Nicht-lineare Komponenten werden durch quadratische Terme in den Modellgleichungen realisiert, da dies zu einfacher nachvollziehbaren Modellen führt. Auch Terme höherer Ordnung wurden betrachtet, allerdings ließ sich mit diesen keine Modellverbesserung mehr erzielen. Als Kriterium für die Verwendung einer Komponente, d.h. für die Beibehaltung eines um sie erweiterten Modells, dient das Informationskriterium von Akaike (AIC). Bei der schrittweisen Erweiterung des Modells wird darauf geachtet, dass sich der Wert dieses Kriteriums nicht erhöht und - wenn möglich - verringert (was meist der Fall war). Diese Vorgehensweise führt tendenziell zum Einschluss von mehr Variablen als es mit konservativeren Kriterien der Fall wäre. Dies ist im vorliegenden Fall auch gewünscht, da das Anliegen die Analyse des vorliegenden Datensatzes in Bezug auf die konkreten Krankenhäuser ist und nicht so sehr die Vorhersage für eine Population von Krankenhäusern. Die beschriebene Strategie führt zu folgendem Modell für die Effektivität (Abbildung 5.8):

```
Therapieerfolg = Baseline
+ logarithmierte Verweildauer
+ quadrierte logarithmierte Verweildauer
+ Aufnahmealter
+ quadriertes Aufnahmealter
+ Aufnahme-GAF
+ Interaktion log. Verweildauer * Aufnahmealter
+ Interaktion quad. log. Verweildauer * quad. Aufnahmealter
+ Geschlecht
+ Zahl der fehlenden Werte
+ Diagnosegruppe
+ Interaktion Schizophrenie * log. Verweildauer
+ Interaktion Schizophrenie * quad. log. Verweildauer
+ Interaktion Affektive Störung * log. Verweildauer
+ Interaktion Affektive Störung * quad. log. Verweildauer
+ Erstaufnahme
+ Rehospitalisierung binnen 30 Tagen
+ Bildung
+ Partnerschafts- und Arbeitsverhältnisse
+ Art der vereinbarten Weiterbetreuung
+ Klinikeffekt
+ zufällige Stationseffekte
+ Fehlerterm
```

Abb. 5.8: Modell zum Vergleich der Effektivität der Krankenhäuser. Random intercept für die Stationen, keine random slopes.

Es enthält quadratische Komponenten für die (logarithmierte) Verweildauer und das Alter zusammen mit linearen und quadratischen Interaktionen zwischen beiden Variablen. Zudem gibt es separate Verweildauereffekte (d.h. Interaktionen) für die Dia-

gnosegruppen „Schizophrenie“ und „affektive Psychosen“. Das Modell zur Kosteneffektivität enthält zusätzlich stationsspezifische Verweildauer-Effekte (random slope für Verweildauer) sowie eine Interaktion zwischen Klinik und Verweildauer. Die Ergebnisse des Krankenhausvergleichs sind, getrennt für Effektivität und Kosteneffektivität, in Tabelle 5.1 zusammengestellt.

	Effektivität		Kosteneffektivität			
	Achsenabschnitt	p-value*	Achsenabschnitt	p-value*	InVD-Effekt	p-value*
KH1	2,3	,04	1,7	,13	,23	,57
KH2	1,0	,4	0,8	,49	,55	,22
KH3	2,5	,08	1,9	,17	,62	,22
KH4	7,1	,00002	6,9	,00003	,014	,98
KH5	3,6	,002	3,4	,003	-,13	,76
KH6	4,7	,002	4,5	,003	-,32	,56
KH7	2,9	,009	2,9	,009	,78	,049
KH8	Referenzhaus					

\*Wald-Test

Tab.5.1: Krankenhausunterschiede in den Modellen zum Effektivitäts- und Kosteneffektivitäts-Vergleich.

Im Effektivitäts-Vergleich (Tabelle 5.1, linke Seite) ergibt sich für die Krankenhäuser eins und vier bis sieben signifikant mehr GAF-Zuwachs, als für das Referenzhaus Nr. acht (das Haus mit dem geringsten adjustierten GAF-Zuwachs). In diesen Häusern werden durchschnittlich zwischen 2.3 und 7,1 GAF-Punkte mehr an Therapieerfolg erzielt, als im Referenzhaus (nach Adjustierung für den unterschiedlichen Casemix). Im Kosteneffektivitäts-Vergleich (Tabelle 5.1, rechte Seite) unterscheidet sich Krankenhaus eins im durchschnittlich erzielten Therapieerfolg (Achsenabschnitt) nicht mehr signifikant vom Referenzhaus Nr. acht, die entsprechenden Ergebnisse für die anderen Krankenhäuser bleiben bestehen. Für den Verweildauer-Effekt (Grenznutzen, durchschnittlicher zusätzlicher Therapieerfolg bei einem zusätzlichen (logarithmierten) Verweildauertag) ergibt sich lediglich für Krankenhaus Nr. sieben ein signifikanter Unterschied zum Referenzhaus Nr. acht. Hier ergibt eine Erhöhung der logarithmierten Verweildauer um eins – entsprechend einer Verlängerung der Verweildauer um den Faktor 2,7 – 0,78 GAF-Punkte mehr an Therapieerfolg, als im Referenzhaus.

## 5.4 Diskussion

Fünf der acht untersuchten Krankenhäuser erscheinen in dieser Analyse als „effektiver“, verglichen mit dem (am wenigsten „effektiven“) Referenzhaus. Berücksichtigt man, dass je Aufenthalt durchschnittlich etwa 15 GAF-Punkte an Therapieerfolg erzielt werden, sind die Krankenhaus-Unterschiede von im Mittel 2,3 bis 7,1 GAF-Punkten teilweise beachtlich. Auch wenn mit 1,6 % nur ein sehr geringer Anteil der Streuung des Therapieerfolgs auf Ebene der Krankenhäuser zu finden ist, ergeben sich doch deutliche Unterschiede zwischen den Häusern. Dabei sind jedoch einige Einschränkungen zu beachten: Zunächst können zu den gefundenen Krankenhaus-Unterschieden auch ein Rater-Bias bzw. unterschiedliche Kodier-Gewohnheiten beigetragen haben. Es könnte etwa eine Rolle spielen, ob der GAF-Score in einem Krankenhaus i.A. bei Aufnahme und Entlassung vom selben Arzt erhoben wird, oder nicht. In diesem Fall wäre beispielsweise eine Vergrößerung des eingeschätzten Therapieerfolgs im Sinne von Kontrasteffekten beim kognitiven Urteil (vgl. etwa Kah-

nemann, Slovic & Tversky, 1982) denkbar. Ein regelmäßiger, auf den GAF-Score gestützter Krankenhausvergleich, der ggf. auch Sanktionen gegen einzelne Häuser begründete, wäre ebenfalls nicht unproblematisch, da der GAF-Score als Rating-Skala auf einfache Weise Manipulationen zugänglich ist. Beispielsweise führte eine systematische Unterschätzung des Aufnahme-GAF bzw. Überschätzung des Entlassungs-GAF zu einem höheren gemessenen „Therapieerfolg“ und entsprechend zu einem besseren Abschneiden in entsprechenden Krankenhausvergleichen.

Wie bei allen Untersuchungen mit nicht-experimentellem Design können auch hier nicht erfasste (hier: Patienten-)Merkmale eine Rolle spielen (hidden variable bias): Beispielsweise könnte der Anteil chronischer Patienten eines Krankenhauses deutliche Auswirkungen auf den durchschnittlich erzielten Therapieerfolg haben, ohne dass dies von den dazu vorhandenen Patientenvariablen (Erstbehandlung ja/nein, Wiederaufnahme innerhalb eines Monats ja/nein) vollständig erfasst sein müsste. Ferner ist zu bemerken, dass der „Zugewinn im Sozialen Funktionsniveau“ (Delta-GAF) zwischen Aufnahme und Entlassung nur ein Aspekt des wohl mehrdimensionalen Konstrukts „Therapieerfolg“ ist. Ebenso wäre neben dem hier betrachteten Therapieerfolg zwischen Aufnahme und Entlassung die Analyse langfristiger Outcomes von großer Bedeutung.

Eine mögliche Erweiterung der vorgestellten Modelle wäre der Einschluss zusätzlicher zufälliger Effekte (random slopes). In den hier betrachteten Modellen wurde lediglich ein Stationseffekt für den durchschnittlichen Therapieerfolg (= Achsenabschnitt, d.h. ein random intercept) bzw. zur Kosteneffektivität noch zusätzlich für den Verweildauer-Einfluss (random slope) berücksichtigt. Es wäre natürlich auch möglich, weitere Effekte wie z.B. den Einfluss des Alters auf den Therapieerfolg stationspezifisch zu modellieren, wie dies etwa in den berichteten Analysen zum Vergleich der Verweildauer gemacht wurde. Allerdings müsste dieser Einfluss dann auch Krankenhaus-spezifisch betrachtet werden, damit nicht tatsächlich vorhandene Krankenhausunterschiede durch Attribuierung auf die entsprechenden Stationen maskiert und somit nicht gefunden werden. (Stationseffekte werden in den Modellen dieses Abschnitts ja lediglich als „Rauschen“ betrachtet, das zwar in den Modellen berücksichtigt werden muss, jedoch nicht weiter interpretiert wird.) Ein Vergleich der Krankenhäuser müsste dann nicht nur Verweildauer-, sondern auch z.B. altersspezifisch erfolgen. Da die Zielsetzung der vorliegenden Analysen ein Overall-Vergleich der betrachteten Krankenhäuser war, wurde darauf verzichtet. Sind die Analysen dagegen eher auf die Unterstützung der Qualitätssicherung in den einzelnen Häusern gerichtet, so erscheint es sinnvoll, Stationseffekte detaillierter zu modellieren. Dazu können, wie in den Analysen zur Verweildauer (vgl. Kap.4), zusätzlich Stationsmerkmale, stationspezifische Effekte von Patientenmerkmalen, sowie auch Interaktionen zwischen Stations- und Patientenmerkmalen in die Betrachtungen miteinbezogen werden. Je nach Zielsetzung der Analyse sind unterschiedliche Arten der Modellierung mit dem sehr flexiblen Instrument hierarchischer Regressionsmodelle möglich.

## **6. Konsequenzen und Ausblick**

Krankenhausvergleiche sind kein „Elfenbeinturm-Thema“ der Versorgungsforschung, im Gegenteil: Die Resultate dieser Studie haben unmittelbare Konsequenzen für die aktuellen gesundheitspolitischen Entscheidungen. Krankenhausvergleiche werden in

der gegenwärtigen gesundheitspolitischen Debatte meist in einem direkten Zusammenhang mit der „Stärkung von Wettbewerbselementen“ genannt, und zumeist nicht als wissenschaftliches Programm, sondern als Hilfsmittel verstanden, um Wettbewerb überhaupt erst auszulösen (vgl. (Lüngen, Wolf-Ostermann, & Lauterbach 2001)). Zum gegenwärtigen Zeitpunkt ist aber nicht absehbar, ob sich die mit Wettbewerb als Steuerungsprinzip im Gesundheitswesen verbundenen Hoffnungen werden einlösen lassen (vgl. (Robra, Swart, & Felder 2002)). Insbesondere für einen „Qualitätswettbewerb“ fehlen auf vielen Feldern noch wirksame Instrumente, wie aus dem bloßen Vergleich dann auch institutionelle Veränderungen erwachsen könnten.

Bislang wurden Krankenhausvergleiche fast ausschließlich unter den Vorzeichen betriebswirtschaftlicher Effizienz (rund um die Festlegung von Kostenerstattungen) oder im Zusammenhang mit Qualitätssicherungsprojekten (als Hilfsmittel für „Benchmarking“) diskutiert. Aber auch in der Diskussion um Qualitätssicherung hat sich der Fokus verbreitert, in dem z.B. Cording für die künftige Orientierung der Qualitätssicherung in der Psychiatrie fordert: *„Ziel muss die Optimierung der psychiatrischen Versorgung unter gesamtgesellschaftlicher Perspektive sein, nicht die betriebswirtschaftliche Optimierung einzelner Institutionen“* ((Cording 2003), S.226).

In der aktuellen, vierten Fortschreibung des Vertrages zwischen den Spitzenverbänden der Krankenkassen und der Deutschen Krankenhausgesellschaft (vom 18. Dezember 2001) über die Krankenhausvergleiche gemäß §5 Bundespflegesatzverordnung wurde ein wichtiger Zweck von Krankenhausvergleichen noch einmal festgeschrieben: Sie sollen die Vertragsparteien bei der „Bemessung von medizinisch leistungsgerechten Budgets und tagesgleichen Pflegesätzen unterstützen“. M.a.W., es geht ums Geld.

*„Der Vergleich von Krankenhäusern bezieht sich auf die krankenhausesbezogenen Vergleichsdaten. Er wird anhand einer Aggregation der ICD- und OPS-Statistiken aller Abteilungen eines Krankenhauses zu einer fiktiven ICD- und OPS-Statistik des Krankenhauses durchgeführt“* (Anhang 2, Abschnitt 2 des Vertrages). Damit werden Krankenhausvergleiche –in der methodischen Sprache unserer Studie formuliert- als reine Level-3-Vergleiche behandelt. Am Beispiel der Verweildauer haben wir hier aufgezeigt, dass die Verweildauer kein Merkmal der Krankenhäuser ist, sondern eine Variable auf Patienten-Ebene. Durch die Mittelwertbildung auf höheren Aggregationsebenen (Station oder Abteilung, Krankenhaus) wird eine formal und inhaltlich falsche Betrachtung eingeführt, die zu völlig falschen Schlüssen führen kann, wie am Beispiel unseres freiwilligen Vergleichs aufgezeigt werden konnte.

Ein Level-3 Vergleich von Häusern führt dazu, dass die Abhängigkeit der Verweildauer von individuellen, Patienten-bezogenen Variablen nicht - wie in einem fairen Vergleich notwendig – ausreichend berücksichtigt werden kann. Die derzeit beschlossene Auswahl von Vergleichskrankenhäusern anhand eines ähnlichen Diagnosenprofils zum betrachteten Krankenhaus kann diese Adjustierung nicht leisten:

1. sind alle Verweildauer-Determinanten außer der reinen ICD-Diagnose nicht berücksichtigt. In unserer Studie an über 27.000 PatientInnen haben das Lebensalter, der Wiederaufnahmestatus, die erreichten therapeutischen Fortschritte beim sozialen Funktionsniveau, die Verträglichkeit der Medikation, und –ganz zentral- die Möglichkeiten zur anschließenden Weiterbehandlung/-versorgung eine eigenständige Be-

deutung für die Verweildauer, die weder über die ICD-Diagnosen, noch durch die Psych-PV-Kategorien angemessen berücksichtigt werden kann.

2. Die Diagnose bedeutet (zumindest hier für die Psychiatrie bei der Diagnose Schizophrenie nachgewiesen) je nach Stations-Kontext, in dem sie behandelt wird, einen unterschiedlich zu berücksichtigenden Einflussfaktor. Kontexteffekte wie sie in unserer Studie für den Anteil an neurotischen PatientInnen und den Anteil an schizophrenen PatientInnen pro Station als Verweildauer-beeinflussend nachgewiesen wurden, werden in der Aggregation auf ein Krankenhaus systematisch ausgeblendet. Allerdings kann es je nach exakter Formulierung des Vergleichszweckes (vgl. Abschnitt 5) auch sinnvoll sein, Kontexteffekte in spezifischen Analysen nicht zu berücksichtigen.

Es bleibt für diese Kritik am abgeschlossenen Vertrag auch ohne Belang, ob die Auswahl von Vergleichshäusern über einen Vektor von Diagnosenprozenten verläuft (wie für die somatischen Krankenhäuser vorgesehen), oder über einen wie auch immer gearteten „ergänzenden Datenkranz“ (wie für die Psychiatrie vereinbart). Denn nicht an der Auswahl der Variablen entscheidet sich diese methodisch orientierte Kritik, sondern an der Auswahl des Vergleichsverfahrens.

Wenn nun aber die (wie hier aufgezeigt) tatsächlich über zumindest drei Datenebenen verlaufende Determination des Verweildauergeschehens nur auf der höchsten Ebene der Krankenhäuser betrachtet wird, erscheint ein „ökologischer Fehlschluss“ bei der Beurteilung von konkreten Krankenhäusern nahezu unvermeidlich. Es werden nicht nur die relevanten Einflussfaktoren nicht oder nicht ausreichend berücksichtigt; vielmehr werden womöglich auch völlig irrelevante, ja irreführende Merkmale der Verweildauer-Variation über die Häuser als sanktionsleitende Ergebnisse produziert.

Um diesen Effekt des „ökologischen Fehlschlusses“ im abgeschlossenen Vertrag nach §5 BPfIV für die Situation in der Psychiatrie genauer zu untersuchen, haben wir neben den hier vorgelegten Analysen mit empirischen Daten zusätzlich eine größere Simulationsstudie unternommen (Frick et al. 2003), auf deren Ergebnisse hier nur kurz hingewiesen werden soll.

Das Rationale dieser Simulationsstudie ist folgendes: Es gebe für alle Psychiatrie-PatientInnen eine exakt bestimmbare, „richtige“ Behandlungsdauer, die ausschließlich von der jeweiligen Diagnose und wenigen, benennbaren anderen Krankheitsvariablen abhängig sei, und bei der keinerlei Unsicherheit über den optimalen Entlassungszeitpunkt existiere. Jede im Vergleich zu dieser feststehenden notwendigen Verweildauer kürzere Behandlung wäre dann eine Unterbehandlung, jede längere Behandlung wäre eine Fehlbelegung. Wenn nun alle ÄrztInnen exakt in diesem optimalen Rahmen behandeln, muss ein methodisch einwandfreier Krankenhausvergleich zu dem Ergebnis kommen, dass trotz der unterschiedlichen Patientenzusammensetzung jedes Krankenhauses keine Unterschiede zwischen den Krankenhäusern bestehen. Denn definitionsgemäß wird in diesem Modell ja jeder Patient in jedem Krankenhaus mit der optimalen Verweildauer behandelt.

Genau diese „optimale Behandlungswelt“ lässt sich in einer Computersimulation unschwer erzeugen. Wir haben uns dabei bemüht, die Randbedingungen der virtuellen

Krankenhauslandschaft möglichst exakt an den realen bundesdeutschen Gegebenheiten auszurichten.

Systematisch variierten wir in unserer Simulationsstudie:

- den Grad der diagnostischen Spezialisierung unter den Leistungsanbietern,
- die Verteilung der Größenklassen der Leistungsanbieter (Bettenzahl),
- die Gesamtzahl der konkurrierenden Anbieter (also Bundesländergröße),
- und den Schwellenwert des im Vertrag fixierten „Ähnlichkeitsmaßes“, ab dem ein Haus für vergleichbar mit einem anderen Haus definiert wird.

Dies sind die wichtigsten Strukturmerkmale, für die wir den Einfluss auf die artifizialen Vergleichsergebnisse des gegenwärtig vereinbarten Mechanismus evaluieren konnten. Die Simulationsergebnisse können in fünf Punkten zusammen gefasst werden:

- 1.) Der vereinbarte Vergleichsmechanismus findet Unterschiede zwischen den zu vergleichenden Krankenhäusern auch in einer Situation, wo alle Patienten absolut korrekt nach dem besten Stand des Wissens behandelt werden und jeglicher Einfluss der Organisationsstruktur auf den eingesetzten Ressourcenaufwand ausgeschlossen wurde.
- 2.) Je nach dem Figur-Grund-Kontrast eines konkret zu vergleichenden Hauses zu seinem Umfeld an „vergleichbaren“ Versorgungsinstitutionen wirkt sich der Vertrag per se „unfair“, d.h. bestimmte Häuser systematisch bevorzugend oder benachteiligend aus.
- 3.) Die Richtung eines Bevorzugungs- oder Benachteiligungsbias ist auf komplexe Weise abhängig von zumindest drei (hier via Simulation nachgewiesenen) Umfeldbedingungen, nämlich Größe des Vergleichsgebiets, Ausmaß der Behandlungsspezialisierung und Marktkonzentration.
- 4.) Zusätzlich wirkt sich der gewählte Cut-off-Wert für die Ähnlichkeit auf die Vergleichsergebnisse uneinheitlich aus: Mit steigender Distanz waren sowohl Verbesserungen im Vergleichsergebnis, wie bei anderen Szenarien für dasselbe Referenzhaus auch Verschlechterungen beobachtbar. Neben einer wechselnden Richtung kristallisierten sich teilweise sogar Veränderungen im Vergleichsergebnis mit kurvilinearere Form heraus.
- 5.) Simple Vorhersagen dieser verschiedenen Verzerrungen vom Muster „je größer das Vergleichsgebiet, desto ungünstiger“ sind nicht möglich. Daher ist auch keine einfache Korrektur des Vergleichsalgorithmus um einen bestimmten „Adjustierungsfaktor“ oder ähnliches konstruierbar.

Als Alternative zu dem vertraglich vereinbarten und die Patienteninteressen vollkommen vernachlässigenden Vorgehen plädieren wir daher dafür, Krankenhausvergleiche mit der hier exemplarisch dargestellten Methodik von Multi-Level-Analysen zu bearbeiten. Dies mag der datentechnisch mühsamere, politisch schwieriger kommunizierbare, und von der Realisation her aufwändigere Weg sein. Wir sehen nach mehr als drei Jahren intensiver Beschäftigung mit den methodischen Aspekten von Institutionsvergleichen hingegen keine gangbare Alternative.

Es kann zudem abschließend festgehalten werden, dass sich Simulationstechniken bei der Evaluation von Auswertungs- und Vergleichsmethoden (wie z.B. im §5-Ver-

trag festgelegt) hervorragend dazu eignen, die komplexen Interaktionen aus mehreren Entscheidungsdimensionen in Kenntnis der möglichen Randbedingungen zu antizipieren. Damit werden unbeabsichtigte Wirkungen vor der konkreten Anwendung eines komplexen Bewertungs- und Entscheidungssystems planbar und ggf. auch veränderbar (Bossel 1992). In der Gesundheitspolitik hat sich dazu auch eine formalisierte Vorgehensweise unter dem Stichwort „Health Technology Assessment“ (Poulsen 1999) eingebürgert, die zukünftig eine immer zentralere Bedeutung für die Gesundheitspolitik erlangen wird (Power, Tunis, & Wagner 1994).

## 7. Literatur:

Angermeyer M.C., Kuhn L. (1988) Gender differences in age at onset of schizophrenia. An overview. *European Archives of Psychiatry and Clinical Neuroscience.*, 237: 6, 351-364.

Babor T.F. (2000) Ecological validity or ecological fallacy? Toward an eco-epidemiology for alcohol studies. *Addiction*, 95: 1, 53-55

Backhaus K., Erichson B., Plinke W., Weiber R. (1996) *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*, 8. Auflage, Berlin: Springer.

Barta, W., Frick, U., Gmel, G., Cording, C. (2003). „Wer liefert mehr Therapieerfolg fürs gleiche Geld?“ – Eine Multilevel-Analyse zum Vergleich der Kosten-Effektivität von acht psychiatrischen Kliniken. In: Pfaff H, Schrappe M, Lauterbach KW, Engelmann U, Halber M (Hrsg.) (2003). *Gesundheitsversorgung und Disease Management. Grundlagen und Anwendungen der Versorgungsforschung*. Bern: Verlag Hans Huber, S.127-35.

Betzler, M. & P. Haun. (1998). Krankenhausvergleich - Status quo und Perspektiven. *Der Chirurg* 69: 1300-1304

Bossel, G. (1992). Modellbildung und Simulation - Konzepte, Verfahren und Modelle zum Verhalten dynamischer Systeme. Braunschweig: Vieweg.

Brecht J, Jenke A. (1989) Diagnosespezifische Krankenhausverweildauern in Schleswig-Holstein. *Soz Praeventivmed*; 34, 227-230

Bryk, A.S. & S.W. Raudenbush. (1992). *Hierarchical Linear Models*. Newbury Park: Sage Publications.

Christiansen C.L., Morris C.N. (1997) Improving the statistical Approach to health care provider profiling. *Annals of Internal Medicine, Part 2*, 127, 764-768.

Cohen J. (1960) A coefficient for agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.

Cording C, Gaebel W, Spengler A, Stieglitz RD, Geiselhart H, John U, Netzold DW, Schönell H, Spindler P, Krischker S. (1995) Die neue psychiatrische Basis-Dokumentation. Eine Empfehlung der DGPPN zur Qualitätssicherung im (teil-) stationären Bereich. *Spektrum der Psychiatrie und Nervenheilkunde* 1995; 24: 3-41

Cording, C. (1997) Quality Assurance in Psychosocial Rehabilitation - An Evolutionary Approach. *ASEAN Journal of Psychiatry* 5 (1997) 3-11

Cording, C.: Basisdokumentation als Grundlage qualitätssichernder Maßnahmen. In: *Qualitätssicherung in der Psychiatrie* (1997). Tropon-Symposium, Bd.XI. Hrsg. von M. Berger und W. Gaebel. Springer, Heidelberg-Berlin-New York-Tokio, 33-51.

Cording C, Kipp J, Kukla R, Kunze H, Saß H. (2001) Der neue Krankenhausvergleich nach §5 BpflV und seine Anwendung in der Psychiatrie. Eine Stellungnahme der DGPPN, der Bundesarbeitsgemeinschaft der Träger Psychiatrischer Krankenhäuser (BAG) und der Aktion Psychisch Kranke (APK). *Nervenarzt*; 72: 63-66.

Cording, C. 2003. Plädoyer für ein neues Paradigma psychiatrischer Qualitätssicherung. *Psychiatrische Praxis* 30, no. 4: 225-229.

DeLong E.R., Peterson E.D., DeLong M., Muhlbaier L.H., Hackett S., Mark D.B. (1997) Comparing Risk-Adjustment Methods for Provider Profiling. *Statistics in Medicine*, 16, 2645-2664

Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) (2001) Memorandum zum Krankenhausvergleich nach §5 BpflV für das Fachgebiet Psychiatrie und Psychotherapie. *Der Nervenarzt*, 72: 65-66

Frick U., Laschat M. (1996) Analyse der Strukturqualität: Am Beispiel der neonatologischen Versorgung. *Gesundheitsökonomie & Qualitätsmanagement*, 1: 1, 12-18.

Frick U., Rehm J., Krischker S., Cording C. (1999) Length of stay in a German psychiatric hospital as a function of patient and organizational characteristics – a multilevel analysis. *Int. J. of Methods in Psychiatric Research*, 8: 3, 146-161

Frick U., Krischker S., Hübner-Liebermann B. (1999) " ... aber die Daten taugen doch nichts!" – Empirische Ansätze zur Überprüfung eines Vorurteils. in: C. Cording (Hg.) *Qualität in der Psychiatrie. Sichtweisen von Klinikern, Krankenkassen, MDK und Krankenhausträgern*. Regensburg: Roderer. S. 111-138

Frick U., Barta W., Binder H. (2001) Fallpauschalen in der stationär-psychiatrischen Versorgung: Empirische Evaluation im Land Salzburg. *Psychiatrische Praxis*, 28: Supl. 1, S55-S62

Frick U. (2002) „Fehlbelegung“ und „Fehlbelegungsprüfung“ in der Psychiatrie. Methodische Anmerkungen zu einer fehlgeleiteten Diskussion. *Psycho* 28 (2), 32-39

Frick U., Binder H., Barta W., Cording C. (2003) "Fair ist, Gleiches mit Gleichem zu vergleichen"? Eine Simulationsstudie zu den Krankenhausvergleichen nach §5 Bundespflegegesetzverordnung. *Gesundheitswesen* 65: 1, 8-18.

Gerste, B. (1996). Bildung von Krankenhausgruppen auf Fallmix-Basis. In *Krankenhaus-Report '96*, ed. M. Arnold und D. Paffrath:115-126. Stuttgart: Gustav Fischer.

Goldstein, H. (1995). *Multilevel Statistical Models*. New York: John Wiley.

Häfner H., Riecher-Rössler A., Hambrecht M., Maurer K, Meissner S., Schmidtke A., et al. (1992) IRAOS: an instrument for the assessment of onset and early course of schizophrenia. *Schizophrenia Research*, 6, 209-223

Häfner H., an der Heiden W., Hambrecht M., Riecher-Rössler A. et al. (1993) Ein Kapitel systematischer Schizophrenieforschung – Die Suche nach kausalen Erklärungen

für den Geschlechtsunterschied im Ersterkrankungsalter. *Nervenarzt*, 64: 11, 706-716.

Häfner H., an der Heiden W., Behrens S., Gattaz W., et al. (1998) Causes and Consequences of the Gender Difference in Age at Onset of Schizophrenia. *Schizophrenia Bulletin*, 24: 1, 99-113.

Hastie, T.J., Tibshirani, R.J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Iezzoni L.I., Ash A.S., Shwartz M., Daley J., Hughes J.S., Mackiernan Y.D. (1996) Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health*, 86: 10, 1379-1387.

Jablensky A., Cole S.W. (1997) Is the earlier age at onset of schizophrenia in males a confounded finding? Results from a cross-cultural investigation. *Brit. J. of Psychiatry*, 170, 234-240.

Kahneman, D., Slovic, P. & Tversky, A. (Hg.) (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge, Cambridge University Press.

Lindamer L.A., Lohr J.B., Harris M.J., Jeste D.V. (1997) Gender, estrogen, and schizophrenia. *Psychopharmacological Bulletin*, 33:2, 221-228

Little R., Rubin D.R. (1987) *Statistical analysis with missing data*. New York: John Wiley

Lüngen, M., K. Wolf-Ostermann, & K. Lauterbach. 2001. *Krankenhausvergleich - Betriebsvergleich nach §5 Bundespflegesatzverordnung*. Stuttgart: Schattauer.

Poulsen, P. 1999. *Health technology assessment and diffusion of health technology*. Odense: Odense University Press.

Power, E., S. Tunis, & J. Wagner. 1994. Technology assessment and public health. *Annual Review of Public Health* 15: 561-597.

Reichsthaler, T. (2001). *Zur Effizienz von Krankenhäusern. Wirtschaftlichkeitsvergleiche auf der Basis statistischer Methoden*. Aachen: Shaker.

Richter D. (1999) Krankenhausbetriebsvergleich für psychiatrische Kliniken – Wie sinnvoll ist ein Vergleich der Behandlungsdauern nach Diagnose und Alter? *Das Gesundheitswesen*, 61 (5): 227-233.

Robra, B.-P., E. Swart, & S. Felder. 2002. Perspektiven des Wettbewerbs im Krankenhaussektor. In *Krankenhaus-Report 2002*, ed. M. Arnold: Schattauer.

Ruttimann UE, Pollack MM. (1996) Variability in duration of stay in pediatric intensive care units: a multiinstitutional study. *Journal of Pediatrics*; 128 (1): 35-44.

Sahlmüller, H. (2001). Vom ICD-Mix zum Fallgruppen-Mix - Ein konzeptioneller Vorschlag zur Weiterentwicklung des Betriebsvergleichs unter Einbezug der DRGs. In

*Krankenhaus-Report 2000*, ed. M. Arnold, M. Litsch und H. Schellschmidt: Kapitel 17.  
Stuttgart: Schattauer.

Schneeweiss, S & O Sangha. (2000). Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich? In *Krankenhausbetriebsvergleiche*, ed. G. Sieben und M. Litsch:1313-147. Berlin: Springer.

Schwarz N, Bless H., (1992) Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. in: *L.L. Martin & A. Tesser (Eds.), The construction of social judgment*, Hillsdale: Erlbaum, pp. 217-245.

Smart R.G., Mann R.E. (2000) The impact of programs for high-risk drinkers on population levels of alcohol problems. *Addiction*, 95:1, 37-51.

Sudman S., Bradburn N.M., Schwarz N. (1996) *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.

Von Korff M, Koepsell T, Curry S, Diehr P (1992) Multi-level analysis in epidemiologic research on health behaviors and outcomes, *American Journal of Epidemiology*, 135: 10, 1077-1082

Winer BJ. *Statistical principles in experimental design*. New York: Mc Graw Hill, 1971<sup>2</sup>

Wittchen, H.-U., Mombour, W. & Fydrich, T. (1997). *Manual zum Strukturierten Klinischen Interview für DSM IV, Achse I und Achse II*. Göttingen: Hogrefe.

Wittchen HU, Saß H, Zaudig M, Koehler K. *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-III-R*, 3.Aufl., Weinheim, Basel: Beltz, 1991, S.32f