Universitätsklinikum Ulm
Klinik für Psychiatrie und Psychotherapie II
Ärztlicher Direktor: Prof. Dr. med. Thomas Becker

Behandlungsergebnis stationärer psychiatrischer Behandlung aus verschiedenen Beurteilerperspektiven

Dissertation zur Erlangung des Doktorgrades der Zahnmedizin
der Medizinischen Fakultät der Universität Ulm

vorgelegt von
Sonja Luka
Lauingen a. D.
2013

Amtierender Dekan: Prof. Dr. Thomas Wirth

1. Berichterstatter: PD Dr. Bernd Puschner

2. Berichterstatter: Prof. Dr. Jörn von Wietersheim

Tag der Promotion: 18.12.2014

Für meine Familie

Inhaltsverzeichnis

Abkürzungsverzeichnis III

1 Einleitung 1

1.1 Entwicklung der Versorgungsforschung 1

1.2 Methoden der Ergebnismessung 4

1.3 Auswahl der Instrumente 5

1.3.1 Ergebnisbereiche 6

1.3.2 Perspektive 7

1.4 Erfassung des Behandlungsergebnisses 12

1.4.1 Untersuchungen aus einer Beurteilerperspektive 12

1.4.2 Untersuchungen aus mehreren Beurteilerperspektiven 14

1.5 Zusammenfassung und Fragestellung 21

2 Material und Methoden 23

2.1 Die EMM-Studie 23

2.2 Beschreibung der Untersuchung 23

2.3 Durchführung der Untersuchung 24

2.4 Teilnehmer 24

2.4.1 Patienten 24

2.4.2 Behandler 25

2.5 Instrumente 25

2.5.1 EB-45 25

2.5.2 GAF 26

2.5.3 HoNOS-D 26

2.6 Deskriptive statistische Methoden 27

2.7 Statistische Methoden zur Analyse der Zusammenhänge zwischen den Perspektiven bei Aufnahme und Entlassung 28

2.7.1 Dichotome Bewertung des psychischen Zustands 28

2.7.2 Korrelationen nach Pearson 29

2.8 Statistische Methoden für die Zusammenhänge zwischen den Perspektiven über die Zeit 29

2.8.1 T-Test 29

2.8.2 „Graphical Modelling“ 29

2.8.2 Klinische Signifikanz 30

3 Ergebnisse ... 34
3.1 Stichprobe ... 34
3.2 Vergleich der Beurteilerperspektiven bei Aufnahme ... 35
3.2.1 Beeinträchtigungsschwere ... 35
3.2.2 Dichotome Bewertung des psychischen Zustands ... 36
3.2.3 Zusammenhänge zwischen den Beurteilerperspektiven ... 39
3.3 Beurteilerperspektiven bei Entlassung ... 40
3.3.1 Beeinträchtigungsschwere ... 40
3.3.2 Dichotome Bewertung des psychischen Zustands ... 40
3.3.3 Zusammenhänge zwischen den Beurteilerperspektiven ... 41
3.4 Vergleich der Beurteilerperspektiven über die Zeit ... 43
3.4.1 Beeinträchtigungsschwere ... 43
3.4.2 Zusammenhänge zwischen den Perspektiven ... 44
3.4.3 Behandlungsergebnis ... 47
4 Diskussion ... 52
4.1 Ergebnisdiskussion ... 52
4.1.1 Vergleich der Beurteilerperspektiven bei Aufnahme und Entlassung ... 52
4.1.2 Vergleich der Beurteilerperspektiven über die Zeit ... 54
4.2 Methodendiskussion ... 58
4.3 Limitationen ... 59
4.4 Stärken der Studie ... 60
4.5 Schlussfolgerung und Ausblick ... 60
5 Zusammenfassung ... 62
6 Literaturverzeichnis ... 64
7 Anhang ... 74
Tabellenverzeichnis ... 74
Abbildungsverzeichnis ... 75
Fragebögen ... 76
8 Danksagungen ... 80
9 Lebenslauf ... 81

Abkürzungsverzeichnis

AOK	Allgemeine Ortskrankenkasse
AMDP	Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie
AUC	Area Under Curve
BADO	Basis Documentation
BASIS	Behaviour and Symptom Identification Scale
BE	Beeinträchtigung
BEZ	Zwischenmenschliche Beziehungen
BDI	Beck Depression Inventory
B-PHQ	Brief Patient Health Questionnaire
BPRS	Brief Psychiatric Rating Scale
CAN	Camberwell Assessment of Needs interview
CANSAS	Camberwell Assessment of Need Short Appraisal Schedule
CGI-I	Clinical Global Impression Scale – Improvement
CSSRI	Client Sociodemographic and Service Receipt Inventory
df	Freiheitsgrad
DGPPN	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde
Diff.	Differenz
$\mathbf{D_{krit}}$	Kritische Differenz
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
EB-45	Ergebnisfragebogen-45
EM	Ergebnis Monitoring
EMM	Ergebnis Monitoring und Management
EQ-5D	EuroQol Group 5-Dimension
FBF	Frankfurter Beschwerdefragebogen
GAF	Global Assessment of Functioning
GAS	Global Assessment Scale
GHQ	General Health Questionnaire
HoNOS	Health of the Nation Outcome Scale

HoNOS-D	Health of the Nation Outcome Scale (deutsche Version)
HRS-D	Hamilton Rating Scale for Depression
ICC	Intraklassen-Korrelationskoeffizient
ICD-10	International Statistical Classification of Diseases and Related Health Problems
KIMTP	Kasahara's Inventory for the Melancholic Type Personality
LOCSS	Location of Community Support Scale
LQoLP	Lancashire Quality of Life Profile
LSP	Life Skills Profile
MANSA	Manchester Short Assessment of Quality of Life
MW	Mittelwert
OQ-45	Outcome Questionnaire-45
p	Signifikanzniveau
PANSS	Positive and Negative Syndrome Scale
QLS	Quality of Life Scale
RCI	Reliable Change Index
SBS	Social Behaviour Scale
SCL-90	Hopkins Symptom Check List-90
SD	Standard Deviation (Standardabweichung)
SF	Soziales Funktionsniveau
SF-36	The Medical Outcomes Study Short Form-36
SIP	Sickness Impact Profile
SOZ	Soziale Integration
SUMD	Scale to Assess Unawareness of Mental Disorder
SY	Symptomatik
SYB	Symptombelastung
TAG	Threshold Assessment Grid
TOT	Total score des EB-45
VE	Verhalten
ZSRS	Zung Self-Rating Scale

1 Einleitung

Das Ergebnis einer Behandlung ist in allen Bereichen der Medizin von zentraler Bedeutung. Denn es vermittelt, ob die gewählte Therapie den gewünschten Effekt erzielt oder nicht. Im psychiatrisch-psychotherapeutischen Bereich stellt die Feststellung des Behandlungsergebnisses eine gewisse Herausforderung dar, weil es hier keine konkret messbaren klinischen Parameter gibt. Mit dieser und anderen Herausforderungen der Ergebnismessung beschäftigt sich die in diesem Bereich entstandene Versorgungsforschung, in dessen Rahmen diese Arbeit erstellt wurde.

Mit Hilfe von drei Fragebögen zur psychischen Beeinträchtigungsschwere der Patienten[1], die zu Behandlungsbeginn und bei Behandlungsende von Behandlern, Patienten und geschulten Ratern ausgefüllt wurden, wird in der vorliegenden Untersuchung das Ergebnis stationärer psychiatrischer Behandlung ermittelt und verglichen.

Die Arbeit ist folgendermaßen gegliedert: In der Einleitung wird zunächst auf die Entstehung der Versorgungsforschung eingegangen. Anschließend werden die Methoden der Ergebnismessung vorgestellt, wobei auf die Methode der Fragebogenerhebung ausführlicher eingegangen wird. Danach folgt ein Überblick über die in der Literatur dokumentierten Untersuchungen zur Erfassung der Behandlungsergebnisse. Die Einleitung endet mit der aus diesem Forschungsstand abgeleiteten Fragestellung. Im nächsten Kapitel wird die Datenerhebung erläutert und die verwendeten statistischen Methoden geschildert. Im Anschluss wird von den damit ermittelten Ergebnissen berichtet, welche im darauf folgenden, letzten Kapitel analysiert und inhaltlich und methodisch diskutiert werden.

1.1 Entwicklung der Versorgungsforschung

In diesem Abschnitt werden der Begriff der Versorgungsforschung und seine Entstehung erläutert. Darauf folgt ein kurzer Überblick über die aktuelle Ausprägung

[1] Aus Gründen der Lesbarkeit wird für alle allgemeingültigen Darstellungen auf die Verwendung der weiblichen Form verzichtet. In solchen Fällen schließt die Verwendung der männlichen Form immer auch alle weiblichen Personen mit ein.

dieses Forschungszweigs in den USA, Großbritannien, Australien und Deutschland.

Vor allem durch steigende Kosten im Gesundheitswesen und darin inbegriffen der psychiatrisch-psychotherapeutischen Versorgung wurde diese zu einer Art Produkt, welches auch den Kräften der Marktwirtschaft unterworfen ist. Die Konsequenz daraus ist ein stärkerer Wettkampf zwischen den Anbietern mit dem Ziel möglichst effizient zu arbeiten und die Kosten zu senken. In der weiteren Entwicklung wurden Begriffe wie Kundenzufriedenheit aus der Wirtschaft übernommen und außerdem ein Qualitätsmanagement eingeführt. Dieses beinhaltet die aktive Teilnahme des Patienten an der Entscheidungsfindung sowie die Erfassung des Behandlungsergebnisses [16]. Da demzufolge nicht nur wirtschaftliche Belange ausschlaggebend sind, sondern auch die Qualität der Versorgung, bildete sich in den letzten 15 Jahren ein Forschungszweig aus, der es sich zum Ziel gesetzt hat, die psychiatrisch-psychotherapeutische Versorgung sowohl ambulant als auch stationär zu verbessern. Hierbei richtet sich der Fokus vermehrt auf den Patienten, indem man versucht mittels inzwischen zahlreicher entwickelter Instrumente, meist in Form von Fragebögen, den Behandlungsverlauf und schließlich das Behandlungsergebnis im klinischen Alltag zu erfassen. Durch die kontinuierliche Feststellung des Behandlungsergebnisses kann der Behandler beurteilen, ob seine Therapie anschlägt, um eventuelle Korrekturen vorzunehmen. Die Zielsetzung ist die Ergebnismessung in der täglichen psychiatrisch-psychotherapeutischen Routineversorgung, in die auch der Patient vermehrt mit eingebunden wird [43, 75].

Weltweite Umsetzung

In den USA liegt das Interesse der Versorgungsforschung vor allen Dingen in der Kostensenkung. Dort wird die Ergebnismessung besonders von den dort ansässigen Versicherungsgesellschaften forciert. Auf der anderen Seite gibt es aber auch Bestrebungen die Qualität der psychiatrisch-psychotherapeutischen Versorgung zu verbessern. Für dieses Ziel engagieren sich hier vor allem Verbraucherorganisationen [74, 75].

In Großbritannien geht der Trend zum einen in Richtung evidenzbasierte Medizin und zum anderen auch in Richtung Kostensenkung. Entstanden ist die Versorgungsforschung dort jedoch aus politischen Bestrebungen. Die zuständigen Be-

hörden benötigten ein Instrument, um die Auswirkungen ihrer Investitionen zu überprüfen. Deswegen wurde ein nationales Institut gegründet, welches Informationen über die Kosteneffektivität medizinischer Behandlungen herausgibt. Außerdem wurden nationale Standards für die psychiatrisch-psychotherapeutische Versorgung entwickelt [74, 75].

In Australien scheinen die Bemühungen im Bereich Ergebnismessung bereits weit vorangeschritten. Dort hat man sich vor allem damit auseinandergesetzt, wie Behandlungsergebnisse erhoben werden können. Daraus wurden anschließend von einer staatlichen Institution Vorschläge für die Verwendung bestimmter Erhebungsinstrumente entwickelt. Um die Verwendbarkeit bzw. den Nutzen der Instrumente zu untersuchen, wurden diese Instrumente unabhängig getestet. Aus diesen Tests wurden dann Empfehlungen für die Ergebnismessung entwickelt. Diese Empfehlungen werden auch schon von einigen psychiatrischen Einrichtungen in der täglichen Routine umgesetzt [74, 75].

In Deutschland bildete die Entstehung der evidenzbasierten Medizin die Grundlage für die empirische Fundierung der klinischen Praxis [86]. Die Bestrebungen sind vielfältig, sowohl die Qualität der psychiatrisch-psychotherapeutischen Versorgung zu verbessern, als auch ihre Kosten zu senken. Unterstützt von der Allgemeinen Ortskrankenkasse (AOK) sollen zum Beispiel für die Erkrankungen Alkoholismus und Depressionen Qualitätsindikatoren entwickelt werden [79], was für die integrierte Versorgung von Schizophrenie Patienten in Kooperation mit der DGPPN (Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde) bereits erfolgt ist. Ein weiteres Projekt der DGPPN will außerdem Qualitätsindikatoren für die Behandlungen der häufigsten psychischen Erkrankungen entwickeln [20]. In Deutschland ist die Qualitätssicherung im psychiatrischen Bereich sogar im fünften Sozialgesetzbuch festgeschrieben. Außerdem ist vom Amt für Qualitätssicherung ein Konsenspapier über die Qualitätsmessung im Gesundheitswesen erarbeitet worden [93].

Während in Australien die Einbeziehung des Einzelnen in seine Behandlung im Fokus steht, geht es in den USA und Großbritannien vor allem um Kosteneffizienz. Inzwischen gibt es jedoch auch hier Ansätze, den Schwerpunkt verstärkt auf die Behandlungsqualität zu legen [74].

Insgesamt bewertet scheinen die Punkte Kostensenkung und Verbesserung der Versorgungsqualität trotz unterschiedlicher nationaler Ausprägung der Versor-

gungsforschung von größter Bedeutung zu sein. In Großbritannien und Deutschland wird die Entwicklung der Versorgungsforschung zusätzlich von der evidenzbasierten Medizin beeinflusst. Des Weiteren wird versucht durch die Testung der unterschiedlichen Verfahren der Versorgungsforschung national einheitliche Empfehlungen bzw. Standards für die Durchführung der Ergebnismessung zu entwickeln, um die Qualitätssicherung und -verbesserung der psychiatrisch-psychotherapeutischen Versorgung zu ermöglichen. Die treibenden Kräfte sind in den USA vor allem die Krankenversicherungen, aber auch Verbraucherorganisationen. In Deutschland sind auch die Krankenversicherungen beteiligt, jedoch sind im Besonderen nationale Institute und psychiatrisch-psychotherapeutische Fachgesellschaften an der Entwicklung der Versorgungsforschung involviert. Ähnlich wie in Deutschland wird auch in Großbritannien und Australien die Forschung in der Ergebnismessung zum großen Teil von staatlicher Seite her gesteuert und vorangetrieben.

1.2 Methoden der Ergebnismessung

In diesem Abschnitt werden die wichtigsten Methoden der Ergebnismessung aufgezählt und es wird ausführlich auf die Datenerhebung mit Fragebögen eingegangen, welche anschließend systematisiert werden.

Zu den Methoden der Ergebnismessung gehören Interviews, welche entweder vom Behandler oder von einer außenstehenden Person mit dem Patienten oder aber auch von einem externen Interviewer mit dem Behandler oder den Angehörigen durchgeführt werden. Ebenso gibt es die Möglichkeit der Verhaltensbeobachtungen oder psychophysiologische Untersuchungen, welche allerdings nur einen geringen Anteil an den angewandten Methoden der Ergebnismessung ausmachen [77]. Der größere Teil der Erhebungen findet stattdessen über Fragebögen statt. Diese sind meistens speziell für den psychiatrisch-psychotherapeutischen Bereich entwickelt worden und liegen für verschiedene Perspektiven und Krankheitsbilder vor [43]. Um den Therapieverlauf messen zu können, geschieht das Ausfüllen dieser Fragebögen in regelmäßigen und kontinuierlichen Abständen. Zur Erleichterung der Eingabe und Vereinfachung der Auswertung findet das Ausfüllen oft am PC statt [43, 75, 77].

Einteilung der Fragebögen

Es wurde bereits eine Vielzahl von Fragebögen entwickelt, was die gezielte Auswahl eines geeigneten erschwert [16]. In der Literatur stößt man deswegen auf diverse Vorschläge die verschiedenen Instrumente zu strukturieren und mittels gemeinsamer Eigenschaften in Kategorien zu ordnen.

Wenn man die Beurteilungsverfahren anhand der *Syndromzugehörigkeit* einteilt [77], so erhält man

a) Verfahren, die weder mit einem bestimmten Syndrom noch mit einer Störungsgruppe einen Zusammenhang aufweisen, wie vor allem Selbstbeurteilungsverfahren, die Befindlichkeiten, Stimmungen und psychische bzw. Allgemeinbeschwerden beinhalten, die bei vielen Störungen auftreten und auch in der Normalpopulation von Bedeutung sein können.
b) syndrombezogene Verfahren (z.B. Depressivität, Angst, Zwang usw.): Dies ist die größte Gruppe, welche sich in ein- und mehrdimensionale Verfahren unterteilen lässt. Mehrdimensional bedeutet in diesem Fall, dass ein Patient meist eine Kombination aus mehreren Störungen aufweist und folglich ein breites Spektrum an Auffälligkeiten erfasst werden sollte. Diese Instrumente sind jedoch zeitaufwendig und dadurch weniger verbreitet.
c) störungsgruppenbezogene Verfahren: Diese wurden im Hinblick auf bestimmte Störungsgruppen konzipiert, wie z.B.: dissoziative Störungen, Depressionen, Schizophrenie usw. [77].

Eine weitere Einteilung kann man vornehmen, indem man die Instrumente nach ihrer *Perspektive* einteilt. So gibt es zum einen Selbstbeurteilungsverfahren aus Patientensicht und zum anderen Fremdbeurteilungsverfahren aus Sicht des Behandlers, der Angehörigen oder anderer Personen [24].

1.3 Auswahl der Instrumente

Im Folgenden sollen die Schwierigkeiten der Ergebnismessung bezogen auf die richtige Auswahl der Instrumente beschrieben werden. Wenn man eine Auswahl unter den zahlreichen Instrumenten treffen möchte, sollte man zunächst festlegen, welche Ergebnisbereiche und welche Perspektive erfasst werden sollen.

Zuerst werden die Überlegungen geschildert, die nötig sind, um die Ergebnisbereiche festzulegen. Anschließend wird von den Problemen, die bei der Wahl der

Perspektive entstehen, berichtet. Hierbei wird zwischen den beeinflussenden Faktoren bei Erhebungen aus der Sicht des Patienten (Selbstbeurteilung) und denen bei Erhebungen aus anderen Perspektiven (Fremdbeurteilung) unterschieden.
Die richtige Auswahl der Instrumente erscheint wichtig, weil es einige Untersuchungen gibt, die belegen, dass verschiedene Instrumente oft auch zu unterschiedlichen Ergebnissen führen [24, 42, 48, 70].

1.3.1 Ergebnisbereiche

Eine erste wichtige Frage ist, welche Bereiche von den Fragebögen abgedeckt werden müssen, um eine Aussage über den Therapieverlauf machen zu können. In der Literatur finden sich folgende übereinstimmende Vorschläge hinsichtlich dieser Bereiche: Psychopathologie/Symptombelastung, Lebensqualität, (Behandlungs-) Zufriedenheit, soziale Funktionsfähigkeit und Bedürfnisse [24, 74, 75].
In einem Review von Slade et al. aus dem Jahre 2002 [75] wurde die Literatur ebenfalls nach Vorschlägen zu den Ergebnisbereichen durchsucht. Dabei wurden 16 Bereiche gefunden, die sich in 7 gemeinsame Kategorien einteilen lassen:

1. Wohlbefinden (v.a. Lebensqualität, nur vom Patienten zu bewerten)
2. Wahrnehmung/Gefühl (v.a. Symptome, aus Patienten- und Behandlersicht)
3. Verhalten (z.B. tägliche Routine, Aktivitäten; aus Patienten- und Behandlersicht)
4. körperliche Gesundheit (Krankheitssymptome, Energie, Vitalität; aus Patienten- und Behandlersicht)
5. zwischenmenschliche Beziehungen (soziale Kompetenzen)
6. Gesellschaft (Sicherheit, Kosten, Aufwand, Versorgung)
7. Behandlungszufriedenheit.

Eine andere Untersuchung [27] ermittelte die Bereiche, die für die Patienten von höchster Bedeutung sind. Dabei schien die Reduzierung depressiver Symptome und Ängste am wichtigsten zu sein. Je nach Art ihrer psychischen Erkrankungen ergaben sich allerdings auch unterschiedliche Bereiche, die den Patienten wichtig waren. Insgesamt wurden hier die Symptombereiche Stimmung, Angst und Furcht als die Bereiche bestimmt, die am besten Veränderungen erfassen.
Häufig wird die Lebensqualität als Ergebniskategorie genannt [16, 24, 39, 67, 69, 74]. Wenn die Lebensqualität allerdings über das Level unerfüllter Bedürfnisse bestimmt wird, kommt man unter Umständen zu unterschiedlichen Ergebnissen je

nach Perspektive, weil Behandler und Patienten die Bedürfnisse unterschiedlich einschätzen [69]. So verglich eine Studie die Einschätzung der Lebensqualität aus Patienten- und Behandlersicht. Es wurden Daten von 37 Schizophreniepatienten erhoben. Die Übereinstimmungen zwischen den beiden Perspektiven wurden mit Cohens Kappa und Pearson Korrelationskoeffizienten untersucht. Das Ergebnis war, dass die Bereiche Symptome und Funktion größere Übereinstimmungen aufwiesen, als die Bereiche physische Gesundheit, soziale Beziehungen und berufliche Aspekte [69]. Aufgrund dieser Unterschiede wird oft empfohlen Daten aus diesen beiden Sichtweisen zu erheben [24, 74].

Zu demselben Ergebnis kommt eine weitere Untersuchung, in der 80 Patienten die größten Veränderungen angeben sollten, die sie durch eine psychotherapeutische Behandlung erfahren hatten. Es wurden 90 Veränderungen angegeben, die mittels hierarchischer Cluster-Analyse untersucht wurden. Dadurch ergaben sich vier Bereiche: verbesserte Symptome, verbessertes Selbstverständnis, verbessertes Selbstbewusstsein und größere Selbstdefinition. Folglich scheint den Patienten eine verbesserte Selbsteinschätzung genauso wichtig zu sein, wie eine verbesserte Symptomatik, ganz im Gegensatz zum Ansatz vieler Therapeuten, die sich eher auf eine reine Symptomreduzierung konzentrieren [11].

1.3.2 Perspektive

Bei der Perspektive gibt es zwei grundlegende Überlegungen, die von Bedeutung sind: erstens, ob eine einzige Perspektive reicht oder ob mehrere nötig sind, und zweitens, welche Perspektive bzw. Perspektiven genutzt werden sollen.

Stieglitz und Mitarbeiter [77] stellten die Behauptung auf, dass eine Korrelation zwischen Fremd- und Selbstbeurteilung nur bis zu einem gewissen Grad erreichbar sei: „Selbst- wie Fremdbeurteilung liegen unterschiedliche Beurteilungsperspektiven zugrunde. Sie können sich daher nicht gegenseitig ersetzen, ihnen kommt somit eher eine komplementäre Funktion zu“ ([77], S. 303).

Dies spiegelt sich auch in der Literatur wider. So gibt es viele Befunde, die dafür sprechen, dass mehrere Perspektiven nötig sind, um das Behandlungsergebnis zu erfassen [24, 39, 42, 75, 77, 82]. An einer Stelle wurde dies damit begründet, dass zwischen den Instrumenten selbst so große Unterschiede bestünden, dass diese den eigentlichen Therapieeffekt überlagerten [42]. An anderer Stelle wurde dies ähnlich wie von Stieglitz et al. damit erklärt, dass die Beurteilung des Behand-

lungsergebnisses vom Blickwinkel des Betrachters abhänge. Es wurden fünf Perspektiven aufgelistet: Patient, Kliniker, Kulturgemeinschaft, Angehörige und die Öffentlichkeit. Da jede dieser Gruppen andere Interessenbereiche habe, die einen Behandlungserfolg ausmachten, komme man auch zu unterschiedlichen Ergebnissen. Die Patienten und deren Angehörige zum Beispiel seien vor allem an einer Verbesserung ihrer Lebensqualität interessiert. Die Behandler hingegen konzentrierten sich mehr auf eine Symptomverbesserung, und die Öffentlichkeit möchte möglichst die Kosten gering halten [39].

Zur Wahl der Perspektive lässt sich sagen, dass besonders oft die Behandler- und Patientensicht empfohlen wird [24, 39, 75, 77, 82]. Es gibt aber auch Aussagen, die noch weitere Perspektiven vorschlagen, so wie z.B. Menschen eines gemeinsamen Kulturraumes oder die allgemeine Öffentlichkeit [39], aber auch Angehörige [24, 39, 82] oder Außenstehende [24].

Weitere Hinweise zur Wahl der Perspektive finden sich im bereits oben erwähnten Review von Slade et al. [75], in welchem die Literatur nach gemeinsamen, sich wiederholenden Prinzipien durchsucht wurde. Hierbei wurden 150 Fachartikel gefunden, von denen sieben Artikel Vorschläge zur Durchführung von Ergebnismessung gaben. In fünf von diesen sieben Artikeln wurde die Empfehlung gegeben, aus mehreren Perspektiven zu messen. Auch hier wurde die Behandler- und Patientensicht am häufigsten genannt [75].

Beeinflussende Faktoren der Selbstbeurteilung

Es finden sich in der Literatur einige Aussagen bezüglich der Faktoren, die sich auf die Selbstbeurteilung auswirken können. Diese lassen sich wohl am besten beschreiben, indem man die Frage untersucht, wovon das Ausfüllen eines Patientenfragebogens abhängt.

So wurde darauf hingewiesen, dass die Patienten bestimmte Voraussetzungen mitbringen müssen, um einen Fragebogen ausfüllen zu können. Zum einen müssen sie ein gewisses intellektuelles Niveau vorweisen und die Fähigkeit zur Selbstreflexion besitzen [77, 81, 84]. Das könnte jedoch mit zunehmender Symptomschwere der psychischen Störung problematisch werden. Es wurde deshalb schon mehrmals postuliert, dass die Diagnose und die Schwere der Erkrankung einen Einfluss auf die Beurteilung haben [16, 42, 71, 77, 81, 88]. Zum anderen könnten ebenso der sozioökonomische Status sowie das Geschlecht eine Rolle spielen

[16]. Ein weiterer Punkt, der einen Einfluss haben könnte, wäre die Motivation und damit die Therapie-Kooperation (Compliance) des Patienten [84].

In einer Studie von Trauer et al. aus dem Jahr 2004 [81] wurde versucht, die Faktoren zu bestimmen, von denen das vollständige Ausfüllen eines Patienten-Fragebogens abhängt. Es wurden die Fragebögen „Health of the Nation Outcome Scale“ (HoNOS [90]), „Life Skills Profile“ (LSP [9]) und „Focus of Care“ [9] verwendet, welche von den Behandlern ausgefüllt wurden, und der Fragebogen „Behaviour and Symptom Identification Scale“ (BASIS [17]), der von ihren Patienten nach deren Zustimmung ausgefüllt wurde. Die Untersuchung fand in vier psychologischen Beratungsstellen im australischen Bundesstaat Victoria statt. Von insgesamt 3351 Patienten füllten 924 (32,4 %) den Fragebogen aus. Das Geschlecht spielte beim Ausfüllverhalten keine Rolle. Das Durchschnittsalter derjenigen, die den BASIS ausfüllten, lag mit 35,9 Jahren signifikant niedriger als das derjenigen, die den Fragebogen nicht ausfüllten (41,8 Jahre). Jüngere Patienten waren also eher bereit die Fragebögen auszufüllen. Vielleicht sind sie eher bereit Neuerungen anzunehmen als die älteren. Nur zwei von 85 Patienten mit organischer Diagnose füllten den BASIS aus. Patienten mit Medikamentenmissbrauch (17,0 %) und Angststörung (15,5 %) machten auch einen vergleichsweise geringen Anteil aus. Der Anteil der Patienten mit affektiven Störungen und Persönlichkeitsstörungen lag bei jeweils um 30 % und derjenige für Patienten mit Psychose (v.a. Schizophrenie) lag bei 39 %. Wahrnehmungsstörungen scheinen also nach dieser Studie einen nachteiligen Einfluss auf die Teilnahme an Ergebnismessung zu haben. Außerdem schwankte der Anteil an teilnehmenden Patienten je nach Institution zwischen 10 % und über 50 %. Somit war der größte Faktor, von dem das Ausfüllen abhing, die Institution, in der die Befragung durchgeführt wurde. Über die Ursachen hierfür wurden von den Autoren einige Vermutungen angestellt. Eventuell befanden sich die Patienten, die nicht an der Befragung teilnahmen, in unfreiwilliger Behandlung, so dass ihre Compliance nicht gegeben war. Vielleicht wurden in den jeweiligen Einrichtungen aber auch nur weniger Patienten von ihrem Behandler um die Teilnahme gebeten, weil dieses in der Behandlungsroutine des Behandlers nicht vorgesehen war. Ein weiterer Punkt könnte gewesen sein, dass die Patienten nicht die kognitiven Voraussetzungen besaßen, um einen Fragebogen auszufüllen und deswegen von ihren Behandlern erst gar nicht dazu eingeladen wurden [81].

In einer früheren Untersuchung, die später im Abschnitt 1.4.2 im Detail geschildert wird, wurde vom gleichen Forscherteam das Behandlungsergebnis aus Sicht von Behandlern und deren Schizophrenie-Patienten verglichen. Beide Parteien füllten hierbei den HoNOS [90] Fragebogen aus. Auch hier war der Anteil derjenigen teilnehmenden Patienten gering, die sich unfreiwillig in Therapie befanden. Die Schwere der Erkrankung hatte allerdings keinen Einfluss auf das Teilnahmeverhalten [82]. Entsprechend zeigte sich in einer anderen Arbeit [84], dass Schizophrenie-Patienten ihren jeweiligen psychischen Zustand ebenso gut einschätzen konnten wie die Patienten einer Kontrollgruppe. Die Glaubwürdigkeit und Auswertbarkeit waren dabei unabhängig von der Symptomschwere. Weitere Erkenntnisse waren, dass die Ergebnisse nicht vom Geschlecht beeinflusst wurden, jedoch vom Alter. Denn die älteren Patienten bewerteten ihren Zustand generell etwas schlechter als die jüngeren. Im Rahmen dieser Untersuchung wurden 120 Patienten mit der Diagnose Schizophrenie (nach: „Diagnostic and Statistical Manual of Mental Disorders“ = DSM IV [2]) in symptomatisch stabilem Zustand untersucht, die aus zwei Universitätskrankenhäusern in Ontario (Kanada) rekrutiert wurden. Die Kontrollgruppe bildeten 32 Patienten mit behandelten Depressionen aus ambulanter Versorgung eines dieser Krankenhäuser.

Zusammenfassend lässt sich feststellen, dass das Geschlecht in diesen Untersuchungen keinen Einfluss auf das Ausfüllverhalten hatte [81, 82, 84]. Die Befunde bezüglich des Einflusses des Alters lassen vermuten, dass ältere Patienten ein anderes Ausfüllverhalten haben könnten [81, 84]. Hinsichtlich der Abhängigkeit von der Compliance stimmen die Studien überein. Es füllten mehr Patienten die Fragebögen aus, wenn ihre Motivation zur Behandlung gegeben war. Deswegen ist es wichtig den Patienten den Nutzen und die Bedeutung der Behandlung sowie der Untersuchung näher zu bringen, um ihre Compliance zu erhöhen. Die Behandler könnten zum Beispiel gezielt darauf geschult werden, ihre Patienten zum Ausfüllen der Fragebögen einzuladen und ihnen zu zeigen, dass man an ihrer Rückmeldung interessiert ist [81, 82]. Zum Einfluss der Diagnose und Schwere der Erkrankung fanden sich widersprüchliche Ergebnisse [81, 82, 84]. In zwei Untersuchungen hatte die Schwere der Erkrankungen bei Schizophrenie-Patienten keinen Einfluss auf das Ausfüllverhalten [82, 84]. Daraus lässt sich schlussfolgern, dass stabile Schizophrenie-Patienten ihren Gesundheitszustand beschreiben und bewerten können und damit durchaus in der Lage sind, aktiv an Ergebnismessung

teilzunehmen. Jedoch zeigte sich in der etwas später veröffentlichen Studie von Trauer et al. [81], dass Wahrnehmungsstörungen sich durchaus nachteilig auf das Ausfüllverhalten auswirken können.

Beeinflussende Faktoren der Fremdbeurteilungsverfahren

Wie bereits weiter oben erwähnt gibt es außer der Selbsteinschätzung der Patienten die Möglichkeit der Bewertung durch Außenstehende. Dies können z.B. die Behandler, Verwandte oder geschulte Studienmitarbeiter sein.

Um die beeinflussenden Faktoren der Fremdbeurteilungsverfahren zu untersuchen, wurde in einer Untersuchung [10] der Fragebogen HoNOS [90] von den Behandlern ausgefüllt. Anschließend wurden 42 dieser Behandler zu ihrer Beurteilung des Fragebogens befragt. Hierbei empfanden 87 % der Befragten das Ausfüllen als zeitraubend. Hinzu kam, dass der Nutzen der Erhebung sehr kritisch bewertet wurde. 30 % fanden die Ergebnismessung moderat nützlich und nur 10 % fanden sie sehr nützlich. Hingegen 60 % waren der Meinung, dass die Datenerhebung wenig oder keinen Nutzen für ihre Arbeit bringt.

Es werden hier also zwei Probleme deutlich: zum einen der erhöhte Zeitaufwand für die Behandler und zum anderen die eventuell mangelnde Compliance der Behandler. Die Behandler waren zum Großteil nicht von der Erhebung überzeugt. In diesem Fall lässt sich das damit begründen, dass in dieser Untersuchung eine PC-Version der HoNOS (HoNOSoft) verwendet wurde, welche in der Anwendung und Auswertung einige Schwierigkeiten mit sich brachte. Weitere mögliche Ursachen für die mangelnde Compliance wären mangelnde Rückmeldung der Ergebnisse an die Behandler, sowie mangelnde Schulung der Behandler in der Anwendung der Verfahren. Die mangelnde Compliance ist eng mit dem erhöhten Zeitaufwand verknüpft, denn die Behandler sind nicht bereit Zeit in etwas zu investieren, dessen Nutzen sie in Frage stellen [74]. Tatsächlich verursacht das Ausfüllen und Analysieren von Erhebungsinstrumenten 10 % mehr Zeitaufwand pro Patient für die Behandler [46].

Ein Training der Anwender, eine Vereinfachung der Verfahren und eine stete Rückmeldung der Ergebnisse sind folglich Möglichkeiten um die Compliance der Behandler in dieser Hinsicht zu erhöhen [75]. Das Schulen der Anwender ist allerdings zunächst wieder mit vermehrtem Zeitaufwand und einem erneuten Kosten-

faktor verbunden [77]. Durch den Einsatz von PC-Versionen der Fragebögen kann die Datenerhebung wiederum beschleunigt werden [75].
Weitere mögliche Einschränkungen bei der Fremdbeurteilung wären, dass unterschiedliche Wissensstände der Bewerter eventuell zu unterschiedlichen Beurteilungen führen, und dass bestimmte Bereiche wie zum Beispiel Befindlichkeiten, Stimmungen und individuelle Bedürfnisse des Patienten von Fremdbeurteilern gar nicht erfasst werden können [74, 75, 77].

1.4 Erfassung des Behandlungsergebnisses

In diesem Abschnitt wird ein Überblick über die in der Literatur beschriebenen Methoden der Ergebnismessung gegeben. Zur groben Systematisierung wurden die Methoden in „Untersuchungen aus einer Beurteilerperspektive“ und „Untersuchungen aus mehreren Beurteilerperspektiven“ unterteilt.

1.4.1 Untersuchungen aus einer Beurteilerperspektive

In der Literatur finden sich vor allem Untersuchungen über den Vergleich verschiedener Instrumente einer Sichtweise. Es wurden u.a. deren Handhabung bzw. Anwenderfreundlichkeit, ihre psychometrischen Eigenschaften und ihre Aussagekraft verglichen [10, 25, 52, 55, 70]. Hierdurch versuchte man unter den zahlreichen verfügbaren Instrumenten das für eine jeweilige Situation beste Instrument zu ermitteln. So wurde z.B. die Reliabilität und Validität der HoNOS [90] untersucht und mit anderen Instrumenten aus Patientensicht verglichen („Social Behaviour Scale“ (SBS [92]), „Brief Psychiatric Rating Scale“ (BPRS [56]), „Global Assessment Scale“ (GAS [19]), „Location of Community Support Scale“ (LOCSS [37]), „General Health Questionnaire“ (GHQ [21]), „The Medical Outcomes Study Short Form-36“ (SF-36 [85])). Hierbei ergab sich, dass die HoNOS einfach und schnell auszufüllen waren und die Reliabilität und Validität generell gut waren, jedoch nicht besser als die der anderen Instrumente [55].
In einer anderen Studie wurden die Patientenfragebögen SF-36 [85] und „EuroQol-5D“ (EQ-5D [8]) in Krankenhäusern getestet und die Patienten zur Handhabung der Fragebögen und ihrer Beurteilung des Nutzens von Ergebnismessung befragt. Das Resultat war, dass die Mehrheit der 392 Patienten es leicht bzw. sehr leicht fand, sowohl SF-36 als auch EQ-5D auszufüllen. Nur 7 % (SF-36) und 4 %

(EQ-5D) gaben an, dass die Instrumente schwer zu verstehen seien. Auf die Frage, wie sie den Nutzen von Ergebnismessung bewerteten, antworteten 219 (57 %) von 385 positiv. Das Geschlecht und das Alter hatten hierbei keinerlei signifikanten Einfluss [52].

Ebenfalls aus Patiensicht wurde das Behandlungsergebnis von 92 Patienten in der öffentlichen psychiatrischen Versorgung von Schweden mit den Instrumenten „Lancashire Quality of Life Profile“ (LQOLP [53], Lebensqualität), „Manchester Short Assessment of Quality of Life“ (MANSA [60], Lebensqualität), „Camberwell Assessment of Needs Interview“ (CAN [58], Bedarfe), „Hopkins Symptom Check List-90“ (SCL-90 [13], Symptombelastung) und einem schwedischen Fragebogen zur Behandlungszufriedenheit [26] zu drei Zeitpunkten erhoben: zu Beginn der Behandlung, bei der 18-Monats-Katamnese und der 6-Jahres-Katamnese. Bei diesem Vergleich ergab sich, dass durchschnittlich 66 %, 57 % bzw. 58 % der Varianz der Bewertungen durch nur einen Faktor erklärt werden können, was als generelle Tendenz der Patienten zu einer mehr oder weniger positiven Bewertung gesehen wurde. Die Autoren kommen zu dem Schluss, dass sich die Ergebnisse von verschiedenen Fragebögen aus Patientensicht überlappen, und sie von einem subjektiven Bewertungsfaktor wie zum Beispiel der jeweiligen Stimmungslage beeinflusst sein könnten [25].

Salvi et al. [70] untersuchten hingegen vier Fragebögen aus Sicht der Behandler auf Gemeinsamkeiten und Eignung für den klinischen Einsatz. Untersucht wurden HoNOS [90], „Global Assessment of Functioning” (GAF [34]), „Camberwell Assessment of Need Short Appraisal Schedule” (CANSAS [58]) und „Threshold Assessment Grid” (TAG [76]). Für insgesamt 264 Patienten aus psychiatrisch-psychotherapeutischen Einrichtungen in London füllten die Behandler jeweils alle vier Instrumente aus. Anschließend wurden Zusammenhänge zwischen den Gesamtsummen der vier Instrumente untersucht. Bei der Korrelationsanalyse fand sich eine geringe Abhängigkeit für den CANSAS Summenwert „total unmet needs“ mit den anderen drei Instrumenten. Eine vorangegangene Hauptkomponentenanalyse ergab eine Hauptkomponente, die als genereller Erkrankungsschwere-Faktor interpretiert wurde. Für diesen Faktor ergab sich die stärkste Korrelation mit der TAG-Summe (0,40) und mit dem GAF-Wert (-0,37). Es gab keinen Hinweis auf grundlegende Übereinstimmungen zwischen den Instrumenten. Die Autoren empfehlen daher HoNOS und CANSAS zu verwenden, wenn man eine detaillierte

Charakterisierung von klinischen und sozialen Bedarfen erhalten möchte. CANSAS habe den Vorteil, dass es zusätzlich Informationen über erfüllte Bedarfe erfasst. Wenn es jedoch um den Erkrankungsgrad ginge, könne man zwischen GAF und TAG auswählen. Der GAF sei jedoch die kürzeste Messmethode und der TAG diejenige mit dem größeren Aussagegehalt. Das Fazit der Autoren: Jedes der vier Instrumente könne empfohlen werden, je nachdem welche Informationen man haben möchte und wie viel Zeit man bereit sei zu investieren [70].

Zusammenfassend lässt sich feststellen, dass bereits viele Instrumente zur Ergebnismessung entwickelt worden sind und sehr viele Studien durchgeführt wurden, die diese entweder einzeln oder im Vergleich zu ähnlichen Instrumenten testeten. Die unterschiedlichen Fragebögen zur gleichen Perspektive decken teilweise ähnliche Bereiche ab und ähneln sich häufig in ihrer psychometrischen Qualität. Sie unterscheiden sich jedoch in ihrem Umfang. In Bezug auf die Durchführung von Ergebnismessung im klinischen Alltag scheinen die Punkte Anwenderfreundlichkeit und Praktikabilität an Bedeutung zu gewinnen.

1.4.2 Untersuchungen aus mehreren Beurteilerperspektiven

Die bisher dargestellten Ergebnisse des Forschungsstands machen deutlich, dass es nur sehr wenige Studien gibt, die sowohl die Patientensicht als auch eine Fremdperspektive erfassen, um diese anschließend auf Gemeinsamkeiten und Differenzen zu untersuchen. Im Folgenden werden diese Studien nun im Detail geschildert. Um einen schnellen Vergleich zwischen den Studien dieses Abschnitts zu ermöglichen, sind diese in Tabelle 1 (S. 15) dargestellt. In dieser Tabelle werden die zentralen Merkmale (wie Stichprobengröße, verwendete Instrumente, Perspektive etc.) der verschiedenen Studien einander gegenübergestellt und die zentralen Ergebnisse knapp zusammengefasst.

Studien mit Studienteilnehmern mit der Diagnose Depressionen

Eine Metaanalyse [42], die sich mit der Frage beschäftigte, ob sich die Einschätzung des Behandlungsergebnisses zwischen Behandlern und ihren depressiven Patienten unterscheidet, wurde in den USA durchgeführt. Im Rahmen dieser Metaanalyse wurden Studien ausgewählt, welche vor und nach der Behandlung

Tabelle 1: Untersuchungen mit mehreren Beurteilerperspektiven

Titel	Autor	Jahr	Ort	n	Diagnose	Instrumente	Perspektive	Ergebnis
Comparison of self-ratings and therapist ratings of outpatients' psychosocial status	Eichenberger[15]	2000	Zürich	80	Schizophrenie, schizotype und wahnhafte Störungen affektive Störungen neurotische Störungen	Interview BPRS	Patienten Kliniker	Patienten schätzten sich positiver ein.
Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison	Lambert [42]	1986	Provo (Utah,USA)	1850	Depressionen	ZSRS BDI HRS-D	Patienten Kliniker	Die Kliniker bewerteten das Ergebnis positiver.
Heterogeneity of outcomes in schizophrenia	Ruggeri [67]	2004	Süd-Verona (Italien)	95	Schizophrenie	BPRS GAF CAN LQoLP	Behandler Patienten	Die Behandler bewerten das Ergebnis schlechter als die Patienten.
Unzureichender subjektiver Behandlungsverlauf bei depressiven Patienten in der haus- und nervenärztlichen Praxis	Schneider [71]	2003	Rheinland, Südbaden, München	165 P 70 B	Depressionen	B-PHQ, Arztdokumenta-tion	Patienten Behandler	Ärzte schätzten den Behandlungserfolg positiver ein als ihre Patienten.
Wertigkeit von Fremd- und Selbstbeurteilungsskalen bei chronisch Schizophrenen	Simhandl [73]	1984	Wien ambulant + stationär	37	Schizophrenie	FBF BPRS AMDP-System	Patienten Behandler	Störungen der Auffassung, der Konzentration, der Merkfähigkeit und des Gedächtnisses korrelierten zwischen FBF und AMDP statistisch signifikant. Vereinzelt statistisch signifikante Korrelationen bei schizophrenen Denkstörungen und affektiver Verflachung.
Concordance between patients and their case managers using the Health of the Nation Outcome Scales(HoNOS)	Trauer [82]	2002	Victoria (Australien)	50	Schizophrenie affektiven Störungen Drogen- bzw. Alkoholabhängigkeit	HoNOS-SR (self-rating), HoNOS	Patient Behandler	Patienten gaben insgesamt negativere Bewertungen ab.
Concordance between self- and observer-ratings on Kasahara's Inventory for the Melancholic Type Personality	Ueki [83]	2002	Münster	44	Depressionen	KIMTP, F-List	Patient Eltern Kinder Partner	Die Patienten schätzten ihre Persönlichkeitseigenschaften positiver ein als Beobachter.
Assessing Health Utilities in schizophrenia. A feasibility study	Voruganti [84]	2000	Ontario (Kanada)	120 (32)	Schizophrenie (Depressionen)	Utility measures, PANSS; SUMD, QLS, SIP, GAF; Gurin's Quality of Life Measure	Patient Außenstehende Pflegekräfte	Bewertung unabhängig von Diagnose, Symptomschwere, Geschlecht. Ältere Patienten schätzten ihren Zustand generell schlechter ein. Patient und Behandler schätzten den Gesundheitszustand ähnlich ein.

Anmerkungen: **AMDP =** Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie; **BDI =** Beck Depression Inventory; **B-PHQ =** Brief Patient Health Questionnaire; **BPRS =** Brief Psychiatric Rating Scale; **CAN =** Camberwell Assessment of Needs interview; **FBF =** Frankfurter Beschwerdefragebogen; **GAF =** Global Assessment of Functioning; **HoNOS =** Health of the Nation Outcome Scale; **HRS-D =** Hamilton Rating Scale for Depression; **KIMTP =** Kasahara's Inventory for the Melancholic Type Personality; **LQoLP =** Lancashire Quality of Life Profile; **PANSS =** Positive and Negative Syndrome Scale; **QLS =** Quality of Life Scale; **SUMD =** Scale to Assess Unawareness of Mental Disorder; **ZSRS =** Zung Self-Rating Scale; **P** = Patientenbewertungen, **B** = Behandlerbewertungen.

mindestens zwei der folgenden Instrumente benutzten: (1) „Zung Self-Rating Scale“ (ZSRS [94], Selbstbeurteilungsverfahren); (2) „Beck Depression Inventory“ (BDI [5], Selbstbeurteilungsverfahren); (3) „Hamilton Rating Scale for Depression“ (HRS-D [23], Fremdbeurteilungsverfahren durch geschulte professionelle Interviewer). Es wurden 36 Studien gefunden, die diese Kriterien erfüllten. Der Stichprobenumfang betrug 1850 Patienten. Für alle drei Skalen wurden jeweils die Effektstärken berechnet. Als Ergebnis stellte sich heraus, dass bei ZSRS und BDI signifikant weniger Veränderung der depressiven Symptomatik als bei HRS-D gemessen wurde. Zwischen ZSRS und BDI ergaben sich keine großen Unterschiede. Die Behandler schätzten also hier den Therapieerfolg positiver ein als die Patienten. Es wurde auch festgestellt, dass die Unterschiede zwischen HRS-D, ZSRS und BDI bei psychosozialer Behandlung (Psychotherapie) größer waren als bei medikamentöser Behandlung [42].

Eine Studie aus Deutschland [71] verglich ebenfalls die Behandler- und Patientensicht von Patienten mit depressiver Symptomatik. Dazu wurde die Beurteilung des Behandlungsverlaufs von ambulant betreuten depressiven Patienten aus Patienten- und Haus- bzw. Facharztsicht untersucht. Die Ärzte kamen auch in dieser Studie zu einer deutlich positiveren Bewertung als ihre Patienten. An der Studie nahmen 43 Haus- und 23 Fachärzte aus drei Regionen (Rheinland, Südbaden und München) teil. Der Stichprobenumfang betrug 70 Patienten. Die Patienten füllten zum ersten Termin und sechs bis acht Wochen danach den Selbstbeurteilungsfragebogen „Brief Patient Health Questionnaire“ (B-PHQ [44]) bezüglich ihrer depressiven Symptomatik aus. Die Ärzte füllten zu Beginn eine standardisierte Basisdokumentation mit Daten zur Soziodemographie, Familiengeschichte, Beratungsanlässen, DSM-IV [2] Stressfaktoren, Diagnose gemäß ICD-10 [89], Schweregrad der Depression und therapeutischen Maßnahmen aus. Die Einschätzung des Behandlungsverlaufs wurde nach sechs bis zehn Wochen in einer Verlaufsdokumentation erhoben und erfasste u. a. die Gesamtbeurteilung der Zustandsänderung, die therapeutische Wirksamkeit und unerwünschte Wirkungen der bisherigen Medikation. Ein Behandlungserfolg wurde anhand der prozentualen Veränderung der Werte im B-PHQ für die zwei Messzeitpunkte ermittelt. Es fanden sich keine relevanten Zusammenhänge zwischen Arzt- und Patienteneinschätzung [71].

Auf der anderen Seite finden sich auch Studien, in denen die Patienten sich tendenziell positiver einschätzten. So wie in einer deutsch-japanischen Studie [83], in welcher allerdings anstelle der Behandlersichtweise die Perspektive von Beobachtern mit derjenigen der Patienten verglichen wurde. Diese Beobachter waren Eltern, Kinder oder Partner der Patienten. Die Patienten hatten auch hier Diagnosen aus dem Bereich der Depressionen (ICD 10: F32 oder F33). Die Stichprobe umfasste 44 Patienten aus ambulanter sowie stationärer Betreuung der Universitätsklinik Münster. Die Daten wurden mithilfe des KIMTP, des „Kasahara's Inventory for the Melancholic Type Personality“ ([35], Selbstbeurteilungs- und Beobachterbögen), erhoben. Die Übereinstimmungen wurden mittels Intraklassen-Korrelationskoeffizienten (ICC [72]) berechnet. Die Übereinstimmung zwischen den Selbst- und den Beobachter-Bewertungen für die mittleren Gesamtwerte war in der Gruppe der endogenen Depressionen signifikant. Die durchschnittliche Gesamtpunktzahl des KIMTP war bei den nicht-endogenen Patienten signifikant höher als bei deren Beobachtern. Die Patienten schätzten folglich ihre melancholischen Persönlichkeitsmerkmale positiver ein als ihre Beobachter [83].

An diesen drei geschilderten Studien nahmen nur Patienten mit Depressionen teil. Diese Studien unterscheiden sich allerdings in ihrer Methodik. Wie für Metaanalysen typisch, gelangte diese nur zu einer so großen Stichprobe, weil viele Einzelstudien zusammengefasst wurden. Das schwächt den Wert des Gesamtergebnisses ab, denn die Qualität der einzelnen Stichproben könnte unterschiedlich sein und damit die Richtigkeit der Daten in Frage stellen [43]. In der deutschen Studie wurden nur ambulant und auch teilweise von Hausärzten und nicht von ausgebildeten Psychologen betreute Patienten betrachtet, die eventuell anders zu bewerten sind als stationär in einer psychiatrisch-psychotherapeutischen Klinik betreute Patienten. Interessant ist, dass diese beiden Studien trotzdem zum gleichen Ergebnis gelangten und damit der deutsch-japanischen Studie widersprechen. Bei dieser Studie ist der Aussagewert aufgrund der sehr geringen Stichprobengröße in Frage zu stellen. Der wichtigste Unterschied ist aber, dass die Patienten hier von nahestehenden Personen bzw. Verwandten bewertet wurden. Die persönliche Mitbetroffenheit und Belastung der Verwandten könnte eine Begründung für deren schlechtere Bewertungen sein.

Studien mit Studienteilnehmern mit der Diagnose Schizophrenie

Desweiteren finden sich Studien, die Patienten mit einer Diagnose aus dem Bereich der Schizophrenie befragten. In einer Untersuchung [82] war das Ergebnis ähnlich wie in den beiden zuerst geschilderten Untersuchungen, denn die Behandler-Bewertungen waren auch hier insgesamt positiver als die der Patienten. Auf die Patientengruppe verteilten sich folgende Diagnosen: Schizophrenie (42 %), affektive Störungen (33,8 %) und Drogen- bzw. Alkoholabhängigkeit (7,7 %). Die Behandler und ihre Patienten füllten den HoNOS Fragebogen aus [90], welcher sowohl als Behandler- wie auch als Patientenversion vorlag. Von 195 eingeladenen Patienten nahmen nur 50 an der Studie teil. Zur Datenanalyse wurden Mittelwert-Vergleiche durchgeführt und ICCs [72] berechnet. Der Mittelwert des Totalscores der Patienten war signifikant höher als der ihrer Behandler. Die Mittelwerte der Patienten waren insbesondere für folgende Items der HoNOS signifikant höher als die ihrer Behandler: kognitive Probleme (4), depressive Stimmung (7), Aktivitäten und Alltag (10) und Anpassungsprobleme (11). Der Mittelwert der Behandler war für das Item 6 (Halluzinationen und Wahnvorstellungen) signifikant höher als der der Patienten. Die ICCs reichten von -0,17 (7: Depressive Stimmung) bis 0,65 (3: Alkohol/Drogenmissbrauch). Der ICC für den Totalscore betrug 0,60. Es lässt sich sagen, dass die Patienten durchweg negativere Bewertungen als die Behandler abgegeben hatten.

Der soeben geschilderten Untersuchung widersprechend existieren jedoch zwei Studien aus dem ähnlichen Zeitraum, die Patienten mit chronischer Schizophrenie untersuchten und feststellten, dass hier die Behandler im Vergleich zu ihren Patienten insgesamt negativere Bewertungen abgaben. So wurden im Universitätskrankenhaus Zürich Patienten (ICD 10: F20, F21-28, F3, F14 [89]) in ambulanter Behandlung untersucht [15]. Die Behandler bewerteten zwei Mal innerhalb von sechs Monaten u. a. den psychosozialen Status (BPRS [56]) ihrer Patienten. 80 von 133 Patienten füllten Selbstbeurteilungsbögen zu diesen zwei Zeitpunkten aus. Anschließend wurden mittels multipler Regressionsanalyse diese beiden Perspektiven verglichen [15].

Zu einem ähnlichen Ergebnis kam man in Verona (Italien), denn auch hier bewerteten die Behandler das Behandlungsergebnis häufiger schlechter als die Patienten. Es wurde von 95 Schizophrenie-Patienten (ICD-10: F20), die sich ein Jahr in Behandlung befanden, das Behandlungsergebnis aus Patienten- und Behandler-

sicht verglichen [66]. Die Daten wurden zu Beginn und dann noch einmal bei einer Drei-Jahres-Nachuntersuchung erhoben. Es kamen die Instrumente BPRS [56], GAF [19], CAN [58] und „Lancashire Quality of Life Profile“ (LQoLP [54]) zum Einsatz. BPRS und GAF wurden von den Behandlern ausgefüllt. Die Patienten hingegen füllten CAN und den LQoLP aus. Bei der 3-Jahres-Katamnese wurde eine Verschlechterung des BPRS-Wertes (Symptombelastung) gemessen. Zwei Drittel (74 %) der Patienten blieben jedoch symptomatisch stabil. Der durchschnittliche GAF-Wert verschlechterte sich von 56,5 auf 53,4 Punkte und 47 % der Patienten wurden hier im zeitlichen Verlauf schlechter bewertet [67].

Diese Ergebnisse zeigen, dass sich auch bei der Diagnosegruppe der Schizophrenie kein einheitliches Bild ergibt, was die Bewertung des Behandlungsergebnisses betrifft. Zwei Studien [15, 66] kommen zu einem ähnlichen Ergebnis, eine [82] weicht davon ab. Schwierig ist, wie auch schon zuvor, die Vergleichbarkeit der Untersuchungen. Hier fällt v.a. die Studie von Trauer et al. [82] auf, in der die Behandler positiver bewerteten. Denn bei dieser war die Stichprobe sehr klein und die Daten wurden nur zu einem Messzeitpunkt erhoben.

Studien mit keinen signifikanten Unterschieden in der Bewertung

Neben denen bereits oben dargestellten Möglichkeiten, dass die Patienten sich schlechter oder besser einschätzten als ihre Behandler bzw. Verwandten, folgen nun noch zwei Beispiele für die dritte Variante, dass keine signifikanten Unterschiede zwischen den Bewertungen von Selbst- und Fremdbeurteilern festzustellen waren.

In der ersten Studie (s.a. Abschnitt 1.3.2), wurden die Patientensicht, die Behandlersicht und als dritte Perspektive die Sichtweise von Außenstehenden verglichen. Die Untersuchung fand unter Zuhilfenahme des sogenannten „Utility Measurements“ [80] statt. Dies ist eine Methode zur Abschätzung der Lebensqualität von chronisch Kranken und beinhaltet zwei Schritte: 1. Die während der Krankheit erfahrenen Gesundheitszustände bestimmen 2. Diesen bestimmte Zahlenwerte zuordnen, welche man „Utilities“ nennt [80]. Diese Zahlenwerte werden von den Patienten auf einer visuellen Analogskala, die von 0 (Tod) bis 100 (perfekt oder der am meisten wünschenswerte Gesundheitszustand) reicht, angezeigt. Die Patienten machten Angaben zu drei Krankheitszuständen, dem momentanen Zustand, dem schlechtesten Zustand seit der Erkrankung und dem angestrebten idealen

Zustand in der Zukunft und wiesen diesen mit Hilfe eines Bewertungsprotokolls „Utilities“ zu. Unabhängige Personen bewerteten die Symptomschwere, Einsicht und Lebensqualität. Die Pflegekräfte ordneten der psychischen Verfassung ihrer Patienten ebenfalls „Utilities“ zu. Die Test-Retest-Reliabilität wurde durch sich einmal wöchentlich wiederholende Intervalle untersucht. Die Übereinstimmungsvalidität der Patientenbewertungen wurde durch verblindete unabhängige Beurteiler mit der Bewertung der Symptomschwere, der Lebensqualität und der „Utility“ bestimmt. Neben dem „Utility Measurement“ kamen noch folgende Instrumente zum Einsatz: „Positive and Negative Syndrome Scale” (PANSS [36]), „Scale to Assess Unawareness of Mental Disorder” (SUMD [1], zur Erfassung der Symptome), GAF [19], „Quality of Life Scale” (QLS [30]), „Sickness Impact Profile” (SIP [6]), „Gurin’s Qualitiy of Life Measure” [22]. Für die statistische Analyse wurden Mittelwerte und Standardabweichungen berechnet und Korrelationen (Pearson) durchgeführt. Wie bereits vorweggenommen ergab sich, dass Patienten und Behandler den Gesundheitszustand ähnlich einschätzten [84].

In der zweiten Studie [73] wurden 37 chronisch schizophrene Patienten (in nichtakuten Phasen) in stationärer und ambulanter Betreuung mit zwei Fremdbeurteilungsverfahren („Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie“, AMDP [28] und BPRS [56]) untersucht. Anschließend wurden die so erhaltenen Werte mit denen aus der Selbstbeurteilungsskala FBF („Frankfurter Beschwerdefragebogen“ = FBF [78], Symptomkatalog subjektiver Beschwerden) verglichen. Anhand des BPRS und AMDP definierten die Studienmitarbeiter das „schizophrene Achsensyndrom nach Berner“ [7]. Hierbei handelt es sich um objektiv beobachtbare Ausdruckssymptome. Es wurden nun die subjektiven Beschwerden aus Patientensicht mit den objektiv erhobenen Symptomen aus der Behandlerperspektive verglichen. FBF erfasste dabei v.a. Störungen der Auffassung, der Konzentration, der Merkfähigkeit und des Gedächtnisses. Diese Symptome korrelierten zwischen FBF und AMDP statistisch signifikant. Die Selbstbeurteilungsskala (FBF) brachte in diesem Fall keine Vorteile gegenüber der Fremdbeurteilung, da beide Perspektiven zu ähnlichen Bewertungen gelangten [73].

1.5 Zusammenfassung und Fragestellung

Bei Betrachtung der vorhandenen Literatur fällt auf, dass der Schwerpunkt der meisten Untersuchungen in der Bewertung der praktischen Handhabung und Auswertbarkeit der verschiedenen Erhebungsinstrumente lag. Hierbei wurden die einzelnen Perspektiven oft isoliert betrachtet ohne zu berücksichtigen, welchen Unterschied verschiedene Sichtweisen für die erhobenen Ergebnisse ausmachen. Die wenigen Studien, die das Behandlungsergebnis aus verschiedenen Perspektiven verglichen, konzentrierten sich bis auf zwei Ausnahmen ausschließlich auf die Patienten- und Behandlersicht. Nur eine der geschilderten Studien hatte Daten aus drei Perspektiven erhoben [84]. Des Weiteren wurde oft nur die Beeinträchtigungsschwere zu einem bestimmten Zeitpunkt gemessen und nicht die Bewertung des Behandlungsergebnisses zwischen den Sichtweisen verglichen. Dies ist fragwürdig, denn um das Behandlungsergebnis zu erfassen sind mindestens zwei Messzeitpunkte nötig. Zur statistischen Analyse der Ergebnisse wurden nur Methoden der klassischen, rein mathematischen Statistik verwendet, die keine Rücksicht auf die klinische Bedeutung der Ergebnisse nahm. Auch waren die Ergebnisse der vorhandenen Studien sehr unterschiedlich, so dass sich kein einheitliches Bild ergab. Wie in Tabelle 1 (S.15) ersichtlich, stimmten drei Studien darin überein, dass die Patienten das Behandlungsergebnis positiver einschätzten [15, 67, 83]. Weitere drei Studien gelangten genau zum gegensätzlichen Schluss, dass sich die Patienten nämlich negativer als die Fremdbeurteiler einschätzten [42, 71, 82]. Zwei Studien kamen zu dem Ergebnis, dass die Bewertungen von Patienten und externen Bewertern eine hohe Übereinstimmung besaßen [73, 84]. Die Bewertungen der Patienten unterschieden sich in sieben von acht Untersuchungen von denen der Fremdrater [15, 42, 67, 71, 73, 82, 83]. Die Verwendung von vielen unterschiedlichen Fragebögen könnte ein Grund für die sehr uneinheitlich gestreuten Ergebnisse der Studien sein. Denn die Fragebögen unterschieden sich sowohl in ihren inhaltlichen Schwerpunkten als auch in ihrem Umfang sowie in ihrer Reliabilität und Validität. Auch war ein Vergleich der Ergebnisse zwischen den Studien nur schwer möglich, da viele verschiedene statistische Verfahren zur Auswertung verwendet wurden. Des Weiteren fiel auf, dass die Art der Diagnose einen Einfluss auf das Ergebnis gehabt haben könnte, denn zwei von den drei Studien, bei denen die Patienten das Ergebnis positiver einschätzten, hatten als Probanden Schi-

zophrenie-Patienten. Im Gegensatz dazu lautete bei ebenfalls zwei von den drei Studien, in denen die Patienten das Ergebnis negativer einschätzten, die Diagnose Depression. Auffallend war auch, dass die Stichprobenumfänge von N=37 bis N=120 (mit Ausnahme der Metaanalyse mit N=1850) eher niedrig ausgefallen waren, deswegen wäre die Durchführung einer Studie mit größerem Stichprobenumfang sinnvoll.

Wenn man nun den Fokus auf den Vergleich der Beeinträchtigungsschwere und des Behandlungsergebnisses legt, um eventuelle Gemeinsamkeiten oder Unterschiede zwischen Patienten- und Behandlereinschätzung festzustellen, so ist es sinnvoll, eine zusätzliche Perspektive hinzuzunehmen, um einen unabhängigen neutralen Blickwinkel zum Vergleich zu erhalten. Des Weiteren muss ein geeignetes statistisches Verfahren ausgewählt werden, das dem rein mathematisch-statistischen Behandlungsergebnis eine klinische Dimension gibt, die für den Patienten von Bedeutung ist, so wie eine Einteilung in Gesundung und Erkrankung. Darüber hinaus sollten Daten zu mindestens zwei Zeitpunkten vorhanden sein, um das Behandlungsergebnis ermitteln zu können.

Fragestellung

Diese Arbeit vergleicht anhand einer Stichprobe von N=294 stationär psychiatrisch-psychotherapeutisch behandelten Patienten Krankheitsschwere und Behandlungsergebnis mit unterschiedlichen Messinstrumenten aus den Perspektiven von Patienten, Behandlern und unabhängigen Beobachtern.

Im Einzelnen werden folgende Fragen beantwortet:

1. Krankheitsschwere: Zusammenhänge zwischen den Beurteilerperspektiven bei Aufnahme und Entlassung (querschnittlich).
2. Behandlungsergebnis: Zusammenhänge zwischen den Beurteilerperspektiven über die Zeit (Prä-Post-Vergleich). Hierfür werden verschiedene Methoden der Veränderungsmessung (statistische und klinische Signifikanz) eingesetzt.

2 Material und Methoden

2.1 Die EMM-Studie

Die vorliegende Arbeit entstand im Rahmen der Studie „Ergebnismonitoring und Ergebnismanagement in der stationären psychiatrischen Versorgung“ (EMM [64]). Diese cluster-randomisierte kontrollierte Studie untersuchte, inwieweit kontinuierliches Ergebnismonitoring und -management die Qualität stationärer psychiatrischer Behandlung verbessern kann. In der EMM-Studie wurde von allen Teilnehmern regelmäßig das Behandlungsergebnis aus Patientensicht erfasst (Ergebnismonitoring) und anschließend an die Behandler und Patienten in der Interventionsgruppe zurückgemeldet (Ergebnismanagement). Die Randomisierung fand auf der Ebene der Behandler statt, die von einer unabhängigen Abteilung (Biometrie der Universität Ulm) auf die Interventions- und Kontrollgruppen verteilt wurden.

Ethikvotum und Datenschutz

Der Studienplan wurde von den Ethikkommissionen der Universität Ulm geprüft und genehmigt. Patienten, die für eine Aufnahme in die Studie in Betracht kamen, wurden über den Untersuchungsablauf, die Freiwilligkeit der Teilnahme, die Vertraulichkeit der erhobenen Daten und das Recht auf jederzeitigen Widerruf der Studienteilnahme ohne Nachteile für ihre Behandlung aufgeklärt und konnten anschließend der Teilnahme schriftlich zustimmen.

2.2 Beschreibung der Untersuchung

Die Untersuchung begann im September 2005 in Günzburg mit der konsekutiven Rekrutierung von 294 Patienten aus der „Klinik für Psychiatrie, Psychotherapie und Psychosomatik“ am dortigen Bezirkskrankenhaus. Im März 2007 endete die Untersuchung.

Das Behandlungsergebnis aus Patientensicht wurde während des stationären Aufenthalts anhand eines Fragebogens am Computer erfasst. Das Behandlungsergebnis wurde über den Grad der Beeinträchtigung gemessen, welcher an zwei

Zeitpunkten von den Patienten sowie unabhängigen Beobachtern und Behandlern bewertet wurde.

2.3 Durchführung der Untersuchung

Im Rahmen der EMM-Studie wurden Daten von Patienten, Behandlern und unabhängigen Beurteilern zu Behandlungsbeginn (T_1) und Behandlungsende (T_2) erfasst. Nachdem die Patienten stationär aufgenommen wurden, besuchte sie ein Mitarbeiter der Studie, um sie über den Ablauf aufzuklären. Nach der Zustimmung des Patienten wurde mit dem „Health of the Nation Outcome Scales“ Fragebogen (T_1: HoNOS [90]) die psychische Beeinträchtigungsschwere des Patienten durch den Studienmitarbeiter erfasst.
Anschließend wurden die Patienten in die Bedienung des Eingaberechners eingewiesen, der sich im Büro des Forschungsteams befand. Alternativ erfolgte die Dateneingabe auf Station, falls es einem Teilnehmer nicht möglich war in das Büro zu kommen. Die Patienten füllten dort den „Ergebnisfragebogen 45“ (EB-45 [41]) zu T_1 und T_2 aus. Die Eingabe der Daten erfolgte mit Hilfe der Software „web-AKQUASI“ [64].
Die Behandler wurden angewiesen, rechtzeitig den Zeitpunkt der Entlassung des Patienten mitzuteilen, damit dann noch einmal die Erhebung der Beeinträchtigungsschwere aus Sicht eines Studienmitarbeiters durchgeführt werden konnte (T_2: HoNOS).
Außerdem wurde zu T_1 und T_2 die Beeinträchtigungsschwere der Patienten durch die Behandler mit der Skala „Global Assessment of Functioning“ (GAF [19]) bewertet.

2.4 Teilnehmer

2.4.1 Patienten

Die Patienten stammten aus der stationären Aufnahme der „Klinik für Psychiatrie, Psychotherapie und Psychosomatik“ am Bezirkskrankenhaus Günzburg. Die Abteilungen für Geriatrie und Suchterkrankungen wurden nicht mit einbezogen. Ebenfalls nicht teilnehmen durften diejenigen, deren kurzfristige Entlassung abzusehen war oder deren Beeinträchtigung als zu schwer eingeschätzt wurde. Ange-

sprochen wurden Patienten im Alter von 18-65 Jahren. Sie sollten die deutsche Sprache verstehen und keine Abhängigkeitsstörung, hirnorganische Störung oder demenzielle Erkrankung vorweisen. Außerdem musste ihre schriftliche Zustimmung zur Teilnahme vorliegen. Für ihren Aufwand wurden die Patienten mit 20 € entschädigt.

2.4.2 Behandler

An der Untersuchung nahmen 43 Ärzte teil. Hierbei fielen auf jeden Behandler im Durchschnitt 11-13 Patienten. Für ihre Teilnahme erhielten sie einen Büchergutschein im Wert von 50 €.
Einschlusskriterium für die Ärzte war Behandlungsverantwortung am BKH Günzburg im Sinne von regelmäßigem Patientenkontakt (Ärzte in Weiterbildung, Fachärzte und Psychologen). Ausgeschlossen wurden Oberärzte, PJ-Studenten und Famulanten.

2.5 Instrumente

Die Basisdaten wurden mit dem „Client Sociodemographic and Service Receipt Inventory“ (CSSRI [65]) erhoben. Die Diagnosen wurden der „Basis Documentation“ (BADO [12]) entnommen. Zur Erfassung der Beeinträchtigungsschwere wurden die Instrumente GAF [19], HoNOS-D [3] und der EB-45 [41] verwendet, die im Folgenden näher erläutert werden.

2.5.1 EB-45

Der „Ergebnisfragebogen-45“ (EB-45) ist ein Selbstbeurteilungsinstrument, das störungsübergreifend die wesentlichen Aspekte psychotherapeutischer Veränderung erfassen soll. Er ist die deutschsprachige Anpassung des „Outcome Questionnaire-45.2“ (OQ-45.2), der in den USA routinemäßig verwendet wird [40]. Der Fragebogen enthält 45 Fragen, die fünffach gestuft beantwortet werden können (nie … fast immer). Er lässt sich in drei Subskalen unterteilen: „Symptombelastung“ (SYB), „zwischenmenschliche Beziehungen“ (BEZ) und „soziale Integration“ (SOZ). Der Summenwert (TOT) misst die allgemeine psychische Beeinträchtigung. Als Signalitems dienen Fragen zu Alkohol- und Drogenmissbrauch, sowie Suizidtendenzen und Gewalttätigkeit am Arbeitsplatz [29, 41]. Bei der Untersu-

chung der konvergenten Validität des EB-45 wurden nur geringe bis mäßige Übereinstimmungen mit denen als Kriterien ausgewählten psychometrischen Instrumenten gemessen. Die größte Übereinstimmung (0,4) fand sich zwischen der Gesamtskala des EB-45 und dem globalen Schwereindex der SCL-90. Beim Vergleich klinischer und nicht-klinischer Stichproben wurden deutliche Unterschiede gemessen, die eine Tauglichkeit zur Diskriminierung von Patienten- und Nicht-Patientengruppen, sowohl auf den Subskalen als auch auf der Gesamtskala, belegen (Kolmogoroff-Smirnov-Z-Test, α = 1 %). Außerdem zeigte das Instrument eine gute Veränderungssensitivität (Wilcoxon-Test, α = 5 %) und Kriteriumsvalidität [29].

2.5.2 GAF

Der „Global Assessment of Functioning“ Fragebogen (GAF) ist die am weitesten verbreitete Methode zur Bewertung des psychopathologischen Status und der sozialen Kompetenz des Patienten durch den Behandler. Er wurde von Endicott et al. 1976 aus der „Global Assessment Scale“ (GAS) [19] entwickelt. Die Bewertung erfolgt auf einer Skala von 0-100, welche in 10er Schritten eingeteilt ist. Ein hoher Wert bedeutet ein besseres Funktionsniveau und ein niedriger Wert ein schlechteres. Der GAF-Wert von 50 ist ein Grenzwert, ab dem professionelle Hilfe angezeigt ist [19, 51]. Der Zusammenhang zwischen Symptombelastung, Diagnose und GAF-Werten wurde mehrfach bestätigt. Der GAF vermittelt jedoch nur wenige oder keine Informationen über die Sozialkompetenz und die Arbeitsfähigkeit eines Patienten bzw. kann keine mittleren Schwankungen erfassen. Eine multiple Regressionsanalyse ergab, dass der soziale oder berufliche Funktionsbereich nur 1 % der Streuung in der GAF Bewertung verursachte [51]. Es zeigte sich für den GAF eine hohe Interrater-Reliabilität (r = 0,76 bis 0,91) und eine mäßige Übereinstimmungsvalidität (r = 0,39) [84].

2.5.3 HoNOS-D

Der Fragebogen „Health of the Nation Outcome Scales-D“ (HoNOS-D) ist die deutsche Version der in England entwickelten HoNOS [90], welche die Gesundheit und soziale Funktionsfähigkeit von Patienten mit psychischen Störungen durch Fremdeinschätzung erfassen. Sie bestehen aus zwölf fünffach skalierten (0: kein Problem … 4: schweres Problem) Items, welche die Bereiche „Verhalten“ (beha-

viour), „Beeinträchtigung“ (impairment), „Symptomatik“ (symptoms) und „soziales Funktionsniveau“ (social) abdecken. Es kann auch ein Gesamtwert (total score) über die ersten zehn Items oder alle zwölf Items bestimmt werden. Dieser repräsentiert die Gesamtbelastungsschwere und bewegt sich zwischen 0 und 48. Die Retest-Reliabiltität scheint zufriedenstellend, denn in einer Studie korrelierte die Mehrzahl der Items signifikant hoch zwischen Beginn und Ende der Behandlung. Außerdem wiesen die einzelnen Items des Fragebogens nur eine geringfügige Korrelation untereinander auf [3, 90]. In einer weiteren Studie wurden die einzelnen Items und der Gesamtwert der HoNOS mit Fremdbeurteilungsskalen aus Behandlersicht korreliert. Hier betrug die durchschnittliche konvergente Korrelation der HoNOS mit den anderen Instrumenten r = 0,44 und bestätigte damit nicht die konvergente Validität. Bei Orrell et al. [55] waren die Zusammenhänge zwischen den Gesamtwerten der HoNOS und dem BPRS jedoch gut und bestätigten damit eine gute konvergente Validität. Die Mehrheit der Items und der Gesamtwert besaßen moderate bis ausreichende Veränderungssensitivität mit Effektgrößen von 0,5 – 1,34. Keine zufriedenstellenden Ergebnisse zeigten sich bei der Faktorvalidität. Aus den 66 möglichen Korrelationen zwischen den zwölf HoNOS-Items fanden sich nur fünf mit einem mittleren Wert von r = 0,3 [4].

2.6 Deskriptive statistische Methoden

Die deskriptive Statistik umfasste Mittelwerte und Standardabweichungen für kontinuierliche Variablen, sowie Häufigkeiten und Prozentangaben für Faktoren. Zur Analyse der Zusammensetzung der Stichprobe wurden für folgende Variablen der Mittelwert oder die Häufigkeit berechnet: Alter, Geschlecht, Schulbildung, Arbeit, höchster erreichter Abschluss, Diagnose, Familienstand, Nationalität und Wohnsituation.

Zur Untersuchung des Zusammenhangs zwischen der Zugehörigkeit zur Gruppe der Dropouts und den Patienten-Merkmalen wurden Testverfahren für unabhängige Stichproben verwendet. Für die Analyse nominaler Daten (Häufigkeiten) wurde der Chi^2-Test herangezogen und für die Analyse kontinuierlicher Daten (Alter) wurde der T-Test verwendet. Anschließend wurden diese Werte auf Signifikanz überprüft (Testung 2-seitig mit einer Irrtumswahrscheinlichkeit von 5 %).

Ebenfalls wurden die Daten der Instrumente GAF, HoNOS und EB-45 deskriptiv untersucht und zu diesem Zweck folgende Variablen berechnet: Mittelwert, Median, Standardabweichung, Minimum, Maximum, Schiefe, Kurtosis und Quartile. Des Weiteren wurden für die Summenwerte dieser Instrumente Histogramme mit den Normalverteilungskurven jeweils für T_1 und T_2 erstellt.

2.7 Statistische Methoden zur Analyse der Zusammenhänge zwischen den Perspektiven bei Aufnahme und Entlassung

2.7.1 Dichotome Bewertung des psychischen Zustands

Einteilung in dysfunktional und funktional durch Verwendung von Cut-Offs

Hierzu müssen Grenzwerte (=Cut-Offs) bestimmt werden, die den Übergang vom funktionalen (gesund) zum dysfunktionalen (krank) Bereich kennzeichnen. Diese Cut-Offs werden aus dem Mittelwert einer Normalpopulation und dem Mittelwert der Patienten mit ihren jeweiligen Standardabweichungen berechnet [33].

$$Cut - Off = \frac{(s_1 - \overline{x}_2) + (s_2 \times \overline{x}_1)}{s_1 + s_2}$$ (nach Jacobson & Truax [33])

$\overline{x}_1$, $\overline{x}_2$ = Mittelwerte der Stichproben (Normstichprobe und Pilotstichprobe)

s_1, s_2 = Standardabweichungen der Stichproben

Der Cut-Off für den EB-45 [41] wurde anhand einer Pilotstudie im Vorfeld ermittelt [64]. Es wurden von 50 Patienten EB-45 Daten erhoben und mit Daten aus der deutschen Bevölkerung verglichen. Dabei ergab sich für den Cutoff (EB-45) = 58,8. Weil für den GAF und den HoNOS keine Bevölkerungsstichproben zur Normierung vorhanden waren, wurden dafür Werte aus der Literatur übernommen (siehe Tabelle 2, S. 32).

Mit Hilfe der Cut-Offs wurde dann die Anzahl derjenigen Patienten berechnet, deren psychischer Gesundheitszustand jeweils zu T_1 und T_2 als funktional bzw. dysfunktional einzustufen war. Zur besseren Veranschaulichung wurde aus diesen Angaben ein Balkendiagramm erstellt.

Sensitivitäten und Spezifitäten

Mit der Einteilung in dysfunktional und funktional lassen sich auch die Sensitivitäten und Spezifitäten für jedes Instrument definieren. Die Spezifität bezeichnet diejenigen Patienten, die als gesund also funktional erkannt wurden. Diese werden

auch als richtig negativ bezeichnet. Als Sensitivität wird der Anteil, der als dysfunktional erkannten Patienten, bezeichnet. Diese werden als richtig positiv bezeichnet [49]. Je nachdem wie der Cut-Off gewählt wird, verschiebt sich das Verhältnis von Sensitivität zu Spezifität. Um zu entscheiden, ob ein Testergebnis richtig oder falsch ist, benötigt man einen Referenztest, der die „Wahrheit“ definiert [87]. In dieser Arbeit wurde jeweils der GAF, der EB-45 oder die HoNOS als Referenz gewählt und die prozentualen Sensitivitäten und Spezifitäten zu T_1 und T_2 berechnet. Die Übereinstimmung zwischen den Anteilen von dysfunktional bzw. funktional eingestuften Patienten durch HoNOS und GAF bzw. HoNOS und EB-45 wurden in Vierfeldertafeln mit den jeweiligen Häufigkeiten dargestellt.

2.7.2 Korrelationen nach Pearson

Mittels bivariater Korrelation nach Pearson und anschließender zweiseitiger Testung auf Signifikanz (Signifikanzniveau $p<0,05$), wurden Zusammenhänge zwischen den Instrumenten und ihren Unterskalen untersucht. Die Korrelationen der Summenwerte wurden in Streudiagrammen visualisiert.

2.8 Statistische Methoden für die Zusammenhänge zwischen den Perspektiven über die Zeit

2.8.1 T-Test

Um zu überprüfen, ob eine signifikante Veränderung der Summenwerte der Instrumente sowie ihrer jeweiligen Unterskalen über den Zeitraum T_1 –T_2 stattgefunden hat (Prä-Post-Vergleiche), wurde der T-Test für gepaarte Stichproben verwendet (Testung 2-seitig mit einer Irrtumswahrscheinlichkeit von 5 %).

2.8.2 „Graphical Modelling“

Um Zusammenhänge zwischen den Instrumenten zu untersuchen, wurde „Graphical Modelling“ [14] angewandt. Hierzu wurde das Programm MIM benutzt (http://www.hypergraph.dk/). Man bezeichnet damit einen Graphen, welcher auf dem Konzept bedingter Unabhängigkeit basiert. Die Variablen werden als Knoten dargestellt. Eine Verbindungslinie zwischen zwei Knoten weist auf einen bedingten (konditionalen) Zusammenhang zwischen diesen zwei Variablen hin [66]. Im Ge-

gensatz dazu werden Variablen in diesem Modell als bedingt unabhängig betrachtet, wenn die Partialkorrelationen zwischen ihnen gleich Null sind. In diesem Fall sind die Variablen im Graphen nicht durch eine Linie verbunden. Um ein geeignetes Modell zu ermitteln, wurden zwei kritische Grenzen für die statistische Signifikanz verwendet (0,01 und 0,05) und mittels schrittweiser Rückwärtsselektion die relevanten Zusammenhänge herausgefiltert [63].

Eine Erweiterung hierzu ist das „Graphical Chain Modell“, in dem auch longitudinale Beziehungen dargestellt werden. Die Variablen werden ihrem jeweiligen Zeitpunkt nach in Blöcken eingeteilt. In jedem Block werden die Variablen untereinander korreliert. Die Variablen eines darauffolgenden Blocks werden mit Hilfe von Regressionen mit den Variablen des vorangegangenen Blocks verbunden. Verbindungen zwischen Variablen desselben Blocks sind non-kausal und werden durch Linien dargestellt und Verbindungen zwischen den Variablen verschiedener Blöcke sind potentiell-kausal und werden durch Pfeile dargestellt [62].

Zunächst wurden die angepassten Partialkorrelationskoeffizienten von EB-45, HoNOS und GAF berechnet. Um relevante Strukturen in dieser Korrelationsmatrix zu bestimmen wurde das zugehörige grafische Block-rekursive (Chain) Modell ermittelt.

2.8.2 Klinische Signifikanz

Um festzustellen, ob und in welchem Maße eine Veränderung der psychischen Beeinträchtigungsschwere stattgefunden hatte, wurde das Konzept der klinischen Signifikanz („clinical significance“) von Jacobson et al. [31] zu Grunde gelegt. Es beschreibt eine Kombination von

A) klinischer Signifikanz und

B) reliabler Veränderung.

A) Klinisch signifikante Veränderung bedeutet, dass ein Patient von einem dysfunktionalen in einen funktionalen psychischen Zustand gelangt. Er kommt also mit einem Problem in eine Therapie und geht ohne dieses Problem und mit normaler Funktion heraus bzw. liegt nach der Behandlung deutlich näher an den Werten der Normalpopulation. Hierzu werden die bereits oben erwähnten Cut-Offs verwendet, die die Grenzen zwischen funktional und dysfunktional festlegen.

B) Mit Hilfe des „Reliable Change Index“ (RCI [31]) kann für eine Einzelperson bestimmt werden, ob eine Veränderung auf einer bestimmten Skala so groß ist,

dass sie mit einer Irrtumswahrscheinlichkeit von 5 % nicht mehr als ein Messfehler des Instrumentes aufzufassen ist. Eine Veränderung gilt als reliabel, wenn der RCI ≥ 1,96 ist. Grob kann man von drei Kategorien ausgehen: Ein Patient kann sich im Prä-Post-Vergleich statistisch signifikant verbessern, verschlechtern oder nicht verändern[47].

$$RCI = \frac{x_2 - x_1}{s_{diff}}$$ (nach: Jacobson & Truax, [33])

x_1 = Wert vor der Therapie

x_2 = Wert nach der Therapie

s_{diff} = Standardfehler der Differenz

FAZIT: Nach der Methode der klinischen Signifikanz von Jacobson et al. [34] ist eine Veränderung klinisch signifikant UND reliabel, wenn die Vorher- und Nachher-Werte eines Erhebungsinstruments zwei Kriterien erfüllen:

1. Klinische Signifikanz: Die Messwerte müssen von einem dysfunktionalen in einen funktionalen Bereich übergegangen sein. Diese beiden Bereiche sind durch einen Cut-Off getrennt.
2. Reliable Veränderung: Die Veränderung muss reliabel sein und nicht durch einen Messfehler oder ein Artefakt bedingt sein. Dafür wurde der „Reliable Change Index“ (RCI) festgelegt. Die Veränderung ist reliabel, wenn der RCI ≥ 1,96 ist.

Mit der Kombination von 1. und 2. lassen sich folgende mögliche Hauptzustände beschreiben:

Gesundung: Cut-Off von dysfunktional zu funktional UND RCI Kriterium erfüllt.

Reliabel verbessert: Cut-Off Kriterium nicht erfüllt, RCI Kriterium erfüllt.

Unverändert: Kein Kriterium wurde erfüllt.

Reliabel verschlechtert: RCI erfüllt, aber verschlechtert.

Erkrankung: Cut-Off von funktional zu dysfunktional, RCI erfüllt [45].

Es ist darauf hinzuweisen, dass die Bezeichnung klinische Signifikanz sowohl das Gesamtkonzept als auch den ersten Teilaspekt der Methode bezeichnet. Im Folgenden ist mit klinischer Signifikanz immer das Gesamtkonzept gemeint.

Die Methode der klinischen Signifikanz [33] wurde in dieser Arbeit zur Ermittlung des Behandlungsergebnisses angewandt. Dazu mussten zunächst die Cut-Offs

und die kritischen Differenzen (=D_{krit}) für jedes Instrument festgelegt werden. Als kritische Differenz bezeichnet man die Differenz zwischen Prä- und Post-Wert, die mindestens erreicht werden muss, damit eine Veränderung reliabel ist („reliable change“).

Für den EB-45 [41] ergab sich anhand der Pilotstudie die kritische Differenz (EB-45) = 17,8 Punkte. Weil für den GAF und den HoNOS keine Bevölkerungsstichproben zur Normierung vorhanden waren, wurden für diese die D_{krit}-Werte aus der Literatur übernommen (siehe Tabelle 2).

Tabelle 2: Cut-Offs und D_{krit}-Werte
Anmerkungen: D_{krit} = kritische Differenz, EB-45 = Ergebnisfragebogen-45, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale.

GAF	HoNOS	EB-45
Cut-Off = 60 [38]	Cut-Off = 11 [57]	Cut-Off= 58,8
D_{krit} = 10 [18]	D_{krit} = 8 [57]	D_{krit} = 17,8

Zunächst wurde mit Hilfe der kritischen Differenzen berechnet, bei welchen Patienten eine reliable Veränderung („reliable change“) vorlag. Anschließend wurden Streudiagramme (Beispiel siehe Abbildung 1) für jedes Instrument angefertigt.

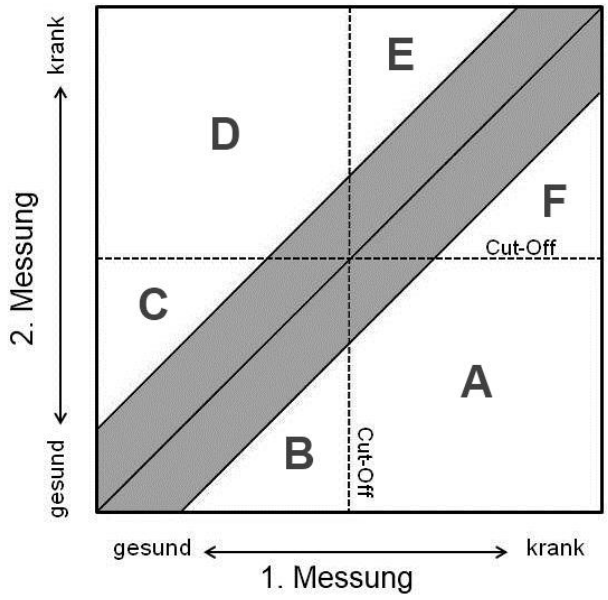

Abbildung 1: Diagramm zur klinischen Signifikanz modifiziert nach [45]
Anmerkungen: A = Gesundung; B+A+F = reliable Verbesserung; C+D+E = reliable Verschlechterung; Band (grau unterlegt) = keine Veränderung.

Bei diesen Streudiagrammen gibt die Abszisse den 1. Messzeitpunkt (T_1) und die Ordinate den 2. Messzeitpunkt (T_2) wieder. Mit Hilfe einer horizontalen und einer vertikalen Linie wurden die jeweiligen Cut-Offs markiert und dadurch die dysfunktionalen von den funktionalen Bereichen getrennt. Zur Veranschaulichung desjenigen Bereiches, in dem keine reliable Veränderung stattgefunden hatte, wurde anschließend innerhalb der Diagramme ein Band (grau unterlegt) festgelegt. Dazu wurde eine Gerade mit der Steigung = 1 durch den Nullpunkt gelegt (auf dieser findet gar keine Veränderung statt) und dann parallel zu ihr links und rechts verschoben zwei Linien im Abstand des Wertes der kritischen Differenz eingezeichnet. Anschließend wurde mit Hilfe von SPSS die Anzahl der Patienten errechnet, für die eine Gesundung, keine Veränderung, reliable Verschlechterung oder reliable Verbesserung angenommen werden konnte. Diese Werte kann man auch aus den Streudiagrammen ablesen.

Die statistische Auswertung wurde mit dem Programm SPSS 11.5.1 für Windows durchgeführt. Die Grafiken wurden ebenfalls mit SPSS erstellt und teilweise mit Microsoft Power Point und dem Bildbearbeitungsprogramm paint.net (http://www.getpaint.net/index.html) modifiziert. Für das „Graphical Modelling“ wurde das Programm MIM verwendet (http://www.hypergraph.dk/).

3 Ergebnisse

3.1 Stichprobe

Von 507 angesprochenen Patienten nahmen zum Zeitpunkt T_1 294 teil. Von 226 Patienten lagen für den EB-45, GAF und die HoNOS Daten zu T_1 und T_2 vor.
Wie in Tabelle 3 (S. 35) ersichtlich, war der Anteil der Frauen etwas geringer als derjenige der Männer. Der jüngste Patient war 17 und der älteste 64 Jahre alt. Knapp zwei Drittel besuchten die Hauptschule, eine Sonderschule oder gaben an keinen Schulabschluss zu besitzen. Ungefähr ein Viertel hatte den Realschulabschluss und ein noch geringerer Anteil das Abitur oder die Fachhochschulreife erreicht. Nicht ganz die Hälfte der Teilnehmer ging einer Vollzeit- oder Teilzeitbeschäftigung nach. Die andere Hälfte der Patienten war ohne Beschäftigung (darunter auch Rentner und Schüler). Über die Hälfte gab als höchsten beruflichen Abschluss die Berufsausbildung an. Deutlich weniger Patienten hatten keinen Abschluss oder eine darunter liegende Qualifikation erlangt. Den geringsten Anteil machten die Patienten aus, die einen Universitätsabschluss besaßen. Die Diagnosen waren vor allem aus der Gruppe der affektiven Störungen (ICD-10: F30-39 [89]). Darauf folgten Schizophrenie, sowie schizotype und wahnhafte Störungen (ICD-10: F20-29). Knapp ein Drittel der Teilnehmer gab an verheiratet zu sein. Etwas mehr waren zum Zeitpunkt der Erhebung ledig und die Übrigen befanden sich in einem anderen Familienstand. Die meisten der Befragten wohnten mit einem Lebenspartner zusammen, die anderen wohnten entweder allein oder gaben eine andere Wohnsituation an.
Die Gruppe von Patienten, von denen für alle drei Skalen nur Daten zu T_1 vorlagen (= Dropouts, N = 68) unterschied sich nicht signifikant von den Teilnehmern zu T_1 und T_2 (Signifikanzniveau $p<0,05$). Es bestand kein signifikanter Zusammenhang zwischen der Zugehörigkeit zur Gruppe der Dropouts und den Stichproben-Variablen.

Tabelle 3: Beschreibung der Stichprobe
Anmerkungen: SD = Standardabweichung, df = Freiheitsgrad, χ2 = Chi-Quadrat, p = Signifikanzniveau; F2, F3, F4, F5, F6 = Diagnosen nach ICD-10.

	T_1 und T_2 (vollständig) N=226	Dropout (nur T_1) N=68	Differenz
Alter			
Jahre, M (SD)	41,96 (11,78)	38,76 (12,32)	$T_{(df=292)}$=1,94; p = ,052
Geschlecht			
weiblich, N (%)	101 (44,7)	38 (55,9)	$\chi^2_{(df=1)}$=2,627; p = ,105
Schulbildung			
Gymnasium, N (%)	33 (14,7)	8 (11,8)	$\chi^2_{(df=2)}$=3,436; p = ,178
Realschule, N (%)	54 (24,0)	24 (35,3)	
Hauptschule, Andere, N (%)	138 (61,3)	36 (52,9)	
Arbeit			
Voll-/Teilzeit, N (%)	99 (43,8)	26 (38,2)	$\chi^2_{(df=1)}$=0,664; p = ,415
Nicht arbeitend, N (%)	127 (56,2)	42 (61,8)	
Höchster erreichter Abschluss			
Universität, N (%)	35 (15,5)	7 (10,3)	$\chi^2_{(df=2)}$=1,593; p = ,451
Berufsausbildung, N (%)	126 (55,8)	43 (63,2)	
Niedriger, N (%)	65 (28,8)	18 (26,5)	
Diagnose			
F2, N (%)	62 (27,4)	23 (33,8)	$\chi^2_{(df=2)}$=1,878; p = ,391
F3, N (%)	128 (56,6)	38 (55,9)	
F4, F5, F6, N (%)	36 (15,9)	7 (10,3)	
Familienstand			
Alleinstehend, N (%)	95 (42,0)	29 (42,6)	$\chi^2_{(df=2)}$=0,736; p = ,692
Verheiratet, N (%)	74 (32,7)	19 (27,9)	
Andere, N (%)	57 (25,2)	20 (29,4)	
Nationalität			
Deutsch, N (%)	206 (91,2)	62 (91,2)	$\chi^2_{(df=1)}$=0,000; p = ,995
Wohnsituation			
Alleine, N (%)	64 (28,4)	22 (32,4)	$\chi^2_{(df=2)}$=1,880; p = ,391
Mit Partner, N (%)	97 (43,1)	23 (33,8)	
Andere, N (%)	64 (28,4)	23 (33,8)	

3.2 Vergleich der Beurteilerperspektiven bei Aufnahme

3.2.1 Beeinträchtigungsschwere

Die zentralen Kennwerte (Mittelwert, Median, Standardabweichung, Minimum, Maximum, Schiefe, Kurtosis, Quartile) für GAF, HoNOS total score und EB-45 TOT zu T_1 sind in Tabelle 4 zusammengefasst. Die Mediane wichen nur geringfügig von den Mittelwerten ab. Es zeigte sich in Übereinstimmung mit Abbildung 2 (S.37), dass die Werte für GAF zum Zeitpunkt T_1 normal und die Werte von HoNOS total score und EB-45 TOT rechtsschief verteilt waren.

Tabelle 4: Kennwerte der Instrumente zu T_1
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, Max. = Maximum, Min. = Minium, MW= Mittelwert, SD = Standardabweichung, T_1 = Zeitpunkt der Aufnahme.

T_1	MW	Median	SD	Min.	Max.	Schiefe	Kurtosis	25 %	50 %	75 %	N
GAF	44,16	45,00	14,996	10	80	-,013	-,411	31,25	45,00	51,00	276
HoNOS total score	11,73	11,00	4,455	2	26	,733	,503	08,00	11,00	14,00	294
EB-45 TOT	79,03	80,00	27,931	11	145	-,172	-,475	57,75	80,00	97,25	294

3.2.2 Dichotome Bewertung des psychischen Zustands

Die absoluten und relativen Häufigkeiten für die Einteilung des psychischen Zustands der Teilnehmer in funktional und dysfunktional sind in Tabelle 5 für HoNOS total score, EB-45 TOT und GAF zu T_1 aufgelistet. Es fällt auf, dass zum Zeitpunkt T_1 beim HoNOS total score (unabhängige Bewerter) fast die Hälfte aller Studienteilnehmer als funktional eingestuft wurde. Beim EB-45 TOT (Patientensichtweise) war dieser Anteil halbiert, so dass nur noch gut ein Viertel als funktional galt. Für die Behandlerseite (GAF) halbierte sich dieser Anteil dann noch einmal auf nur noch 12 Prozent.

Tabelle 5: Einteilung in dysfunktional und funktional zu T_1
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme.

T_1	N	Funktional N (%)	Dysfunktional N (%)
HoNOS total score	294	128 (43,5)	166 (56,5)
EB-45 TOT	294	75 (25,5)	219 (74,5)
GAF	276	33 (12,0)	243 (88,0)

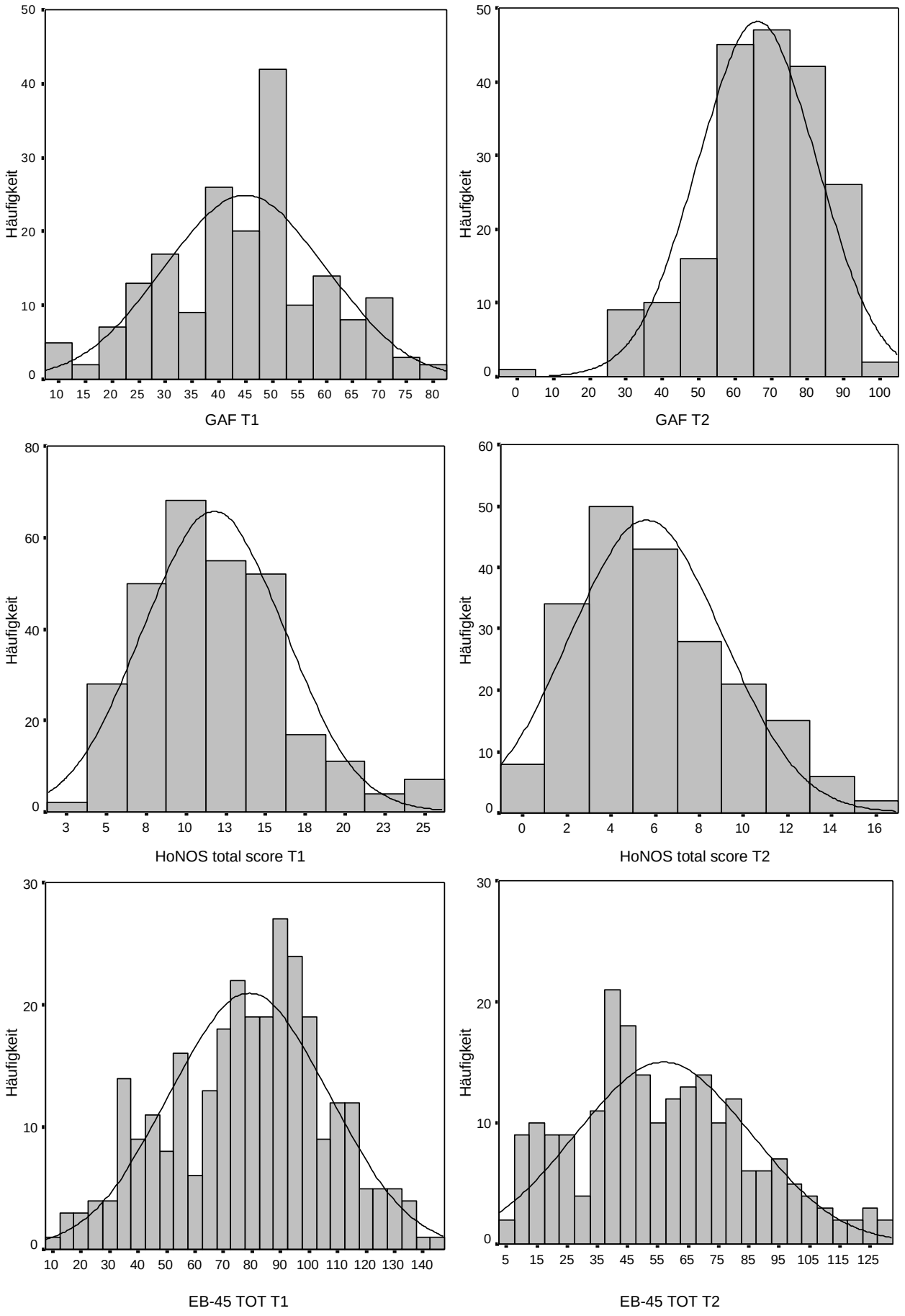

Abbildung 2: Histogramme zu den Instrumenten mit Normalverteilungskurven
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T1 = Zeitpunkt der Aufnahme, T2 = Zeitpunkt der Entlassung.

Die Abbildung 3 zeigt den Grad der Übereinstimmung (Spezifität und Sensitivität) zwischen den drei Beurteilerperspektiven zum Zeitpunkt der Aufnahme (T_1). Es wurden zwei Drittel der Patienten sowohl von den Behandlern als auch von sich selbst als dysfunktional eingestuft. Nur drei Prozent wurden übereinstimmend als funktional bewertet. 69 % (66 % + 3 %) wurden also von beiden Seiten gleich beurteilt. Demgegenüber wurden 31 % (22 % + 9 %) unterschiedlich eingeschätzt.

Etwas geringer als zwischen Behandlern und Patienten war die Übereinstimmung zwischen Behandlern und Außenstehenden, aber sie betrug aber auch fast zwei Drittel. Etwas mehr als die Hälfte der Patienten wurde von beiden als dysfunktional und nur ein sehr geringer Anteil als funktional beurteilt. Zu unterschiedlichen Ergebnissen gelangten sie bei gut einem Drittel der Patienten.

Ebenfalls nicht ganz zwei Drittel der Patienten wurden sowohl von den Außenstehenden als auch von sich selbst als dysfunktional bzw. funktional eingestuft, während etwas mehr als ein Drittel unterschiedlich bewertet wurde.

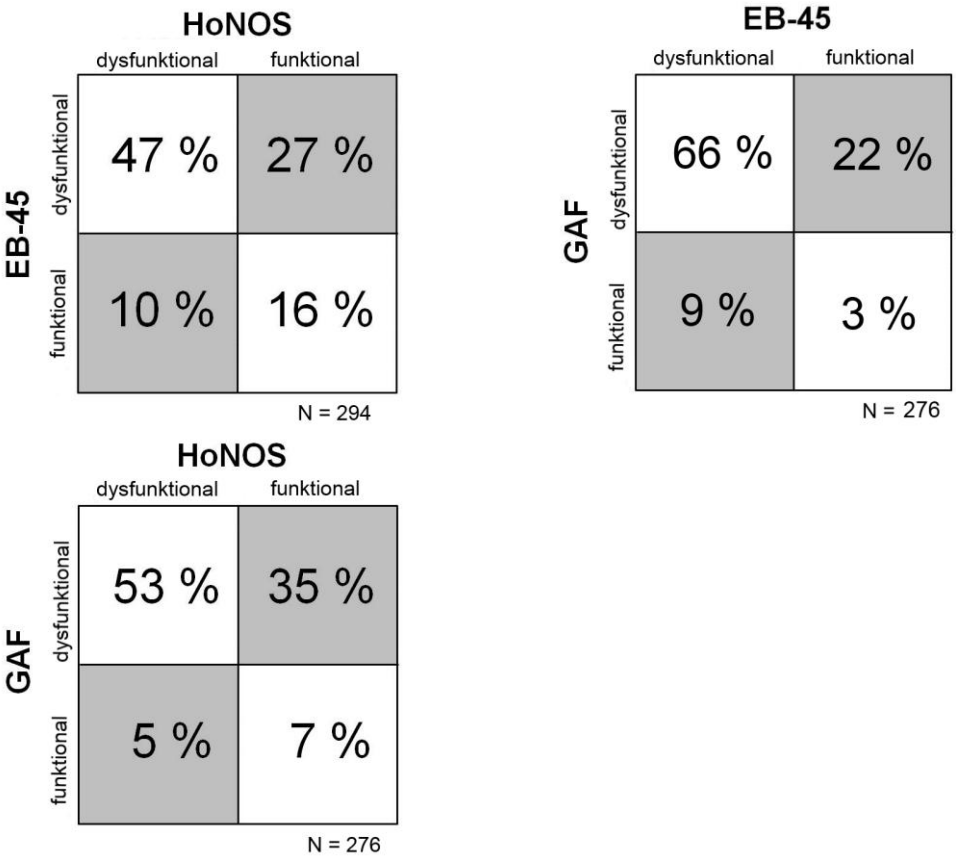

Abbildung 3: Vergleich der Bewertungen des psychischen Zustands zwischen Behandlern, Unabhängigen und Patienten (Spezifität/ Sensitivität) zu T_1

Anmerkungen: EB-45 = Ergebnisfragebogen-45, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme.

3.2.3 Zusammenhänge zwischen den Beurteilerperspektiven

Zur Analyse der Zusammenhänge zwischen den Perspektiven wurden Korrelationen nach Pearson durchgeführt, welche in Tabelle 6 für T_1 angegeben sind. Bei einem Signifikanzniveau von p<0,01 ergaben sich zum Zeitpunkt T_1 vor allem signifikante Korrelationen innerhalb der Unterskalen der HoNOS (Ausnahme „Verhalten“) und innerhalb der Unterskalen des EB-45, sowie zwischen den Unterskalen und dem jeweils zugehörigen Gesamtwert eines Instruments.

Tabelle 6: Korrelationen (Pearson) der Instrumente zu T_1
Anmerkungen: **. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant; *. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant; BE = Beeinträchtigung; BEZ = Zwischenmenschliche Beziehungen; EB-45 TOT = Ergebnisfragebogen-45 total score; GAF = Global Assessment of Functioning; HoNOS = Health of the Nation Outcome Scale; SF= Soziales Funktionsniveau; SOZ = Soziale Integration; SY = Symptomatik; SYB = Symptombelastung; VE = Verhalten, T_1 = Zeitpunkt der Aufnahme.

T_1		**GAF**	**EB-45 TOT**				**HoNOS total score**				
				SYB	**BEZ**	**SOZ**		**VE**	**BE**	**SY**	**SF**
GAF		1									
EB-45 TOT		-,025 p=,684	1								
	SYB	,034 p=,569	,959** p<,000	1							
	BEZ	-,062 p=,304	,752** p<,001	,572** p<,001	1						
	SOZ	,086 p=,152	,825** p<,001	,718** p<,001	,553** p<,001	1					
HoNOS total score		-,187* p=,002	,375** p<,001	,358** p<,001	,295** p<,001	,299** p<,001	1				
	VE	-,082 p=,174	,123* p=,034	,077 p=,189	,208** p<,001	,091 p=,120	,491** p<,001	1			
	BE	-,136* p=,539	,168** p=,004	,206** p<,001	,017 p=,769	,131* p=,025	,535** p<,001	-,036 p=,539	1		
	SY	-,063 p=,547	,414** p<,001	,431** p<,001	,237** p<,001	,324** p<,001	,592** p<,001	,035 p=,547	,174** p=,003	1	
	SF	-,145* p=,012	,216** p<,001	,179** p<,002	,229** p<,001	,188** p=,001	,752** p<,001	,147* p=,012	,281** p=,001	,240** p<,001	1

HoNOS total score und EB-45 TOT korrelierten ebenfalls signifikant. Alle Unterskalen dieser beiden korrelierten signifikant, bis auf „Verhalten“ (HoNOS) mit „Symptombelastung“ (EB-45) und „Verhalten“ mit „soziale Integration“ (EB-45),

sowie „Beeinträchtigung“ (HoNOS) mit „zwischenmenschliche Beziehungen“ (EB-45). Für GAF und HoNOS total score und die Unterskalen „Beeinträchtigung“ und „soziales Funktionsniveau“ ergaben sich ebenfalls signifikante Zusammenhänge. GAF und EB-45 TOT und Subskalen korrelierten nicht signifikant.

3.3 Beurteilerperspektiven bei Entlassung

3.3.1 Beeinträchtigungsschwere

Die zentralen Kennwerte (Mittelwert, Median, Standardabweichung, Minimum, Maximum, Schiefe, Kurtosis, Quartile) für GAF, HoNOS total score und EB-45 TOT zu T_2 sind in Tabelle 7 zusammengefasst.

Betrachtet man die Histogramme (Abbildung 2, S. 37) und die Werte für die Schiefe zum Zeitpunkt T_2 sieht man für den GAF eine linksschiefe Verteilung und für HoNOS total score und EB-45 TOT eine rechtschiefe Verteilung.

Tabelle 7: Kennwerte der Instrumente für T_2
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, Max. = Maximum, Min. = Minium, MW= Mittelwert, SD = Standardabweichung, T_2 = Zeitpunkt der Entlassung.

T_2	MW	Median	SD	Min.	Max.	Schiefe	Kurtosis	25 %	50 %	75 %	N
GAF	66,17	68,50	16,368	0	100	-,751	,906	59,75	68,50	80,00	198
HoNOS total score	5,60	5,00	3,455	0	16	,576	-,313	03,00	05,00	08,00	207
EB 45 TOT	56,81	53,00	28,897	4	128	,371	-,394	37,75	53,00	75,25	218

3.3.2 Dichotome Bewertung des psychischen Zustands

Die absoluten und relativen Häufigkeiten für die in funktional und dysfunktional eingeteilten Teilnehmer sind in Tabelle 8 für HoNOS total score, EB-45 TOT und GAF zu T_2 aufgelistet. Zu T_2 ergaben sich für den HoNOS total score die meisten als funktional bewerteten Teilnehmer, dann folgten der GAF mit knapp zwei Dritteln und der EB-45 TOT mit etwas mehr als der Hälfte an Patienten mit funktionalem psychischen Zustand.

Tabelle 8: Einteilung in dysfunktional und funktional zu T_2
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_2 = Zeitpunkt der Entlassung.

T_2	N	Funktional N (%)	Dysfunktional N (%)
HoNOS total score	214	184 (86,0)	30 (14,0)
EB-45 TOT	218	118 (54,1)	100 (45,9)
GAF	198	126 (63,6)	72 (36,4)

Wie in Abbildung 4 zu sehen, betrug die Übereinstimmung bei der Einteilung in funktional und dysfunktional zwischen Patienten und Behandlern 57 %. Die Übereinstimmung von GAF mit HoNOS total score und EB-45 TOT mit HoNOS total score betrug nun 63 bzw. 62 Prozent. Zu T_2 wurde gut ein Viertel der Patienten von den Behandlern als funktional eingestuft, obwohl sich diese selber als dysfunktional bewerteten.

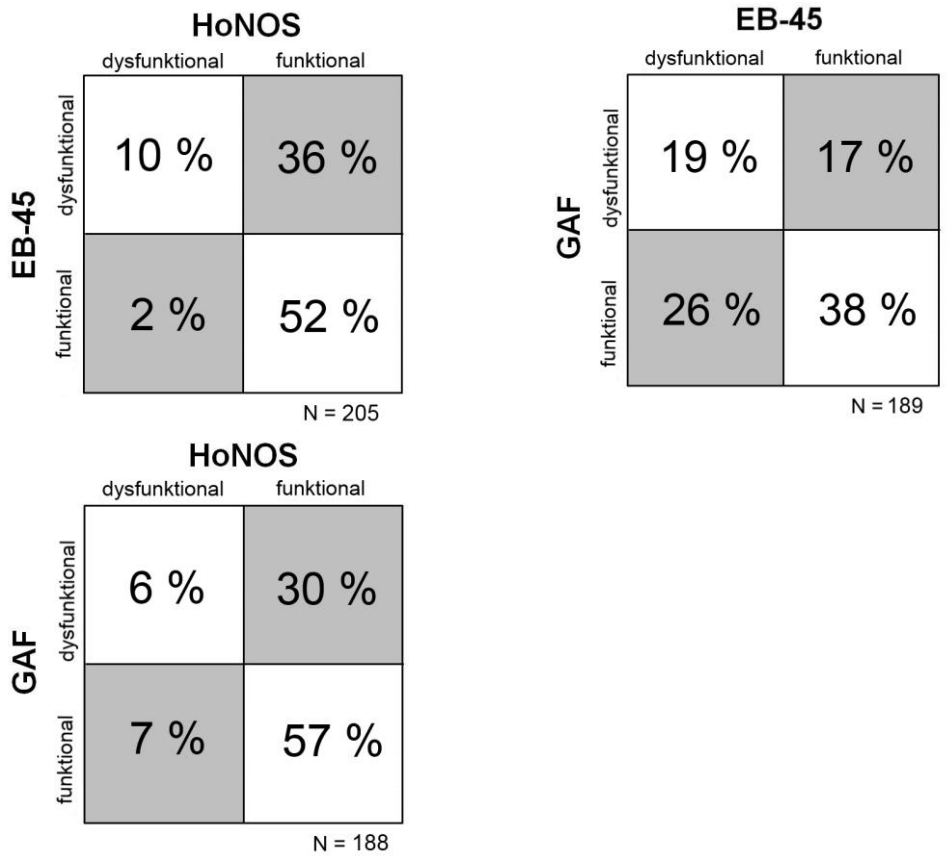

Abbildung 4: Vergleich der Einschätzungen von Behandlern, Unabhängigen und Patienten (Spezifität/ Sensitivität) zu T_2
Anmerkungen: EB-45 = Ergebnisfragebogen-45, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_2 =Zeitpunkt der Entlassung.

3.3.3 Zusammenhänge zwischen den Beurteilerperspektiven

Tabelle 9 zeigt die Ergebnisse der Korrelationen für die Instrumente inklusive ihrer Unterskalen zum Zeitpunkt T_2. Es ergaben sich signifikante Korrelationen innerhalb der Unterskalen des HoNOS und innerhalb der Unterskalen des EB-45, sowie zwischen den Unterskalen und dem jeweiligen Summenwert des Instruments. HoNOS total score und EB-45 TOT korrelierten ebenfalls signifikant.

Tabelle 9: Korrelationen (Pearson) der Instrumente zu T2
Anmerkungen: **. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant; *. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant; BE = Beeinträchtigung; BEZ = Zwischenmenschliche Beziehungen; EB-45 TOT = Ergebnisfragebogen-45 total score; GAF = Global Assessment of Functioning; HoNOS = Health of the Nation Outcome Scale; SF= Soziales Funktionsniveau; SOZ = Soziale Integration; SY = Symptomatik; SYB = Symptombelastung; VE = Verhalten.

T_2		GAF	EB-45 TOT	SYB	BEZ	SOZ	HoNOS total score	VE	BE	SY	SF
GAF		1									
EB-45 TOT		-,044	1								
		p=,546									
	SYB	-,054	,972**	1							
		p=,457	p<,001								
	BEZ	-,044	,802**	,672**	1						
		p=,552	p<,001	p<,001							
	SOZ	,010	,788**	,788**	,618**	1					
		p=,897	p<,001	p<,001	p=,000						
HoNOS total score		-,180*	,521**	,513**	,429**	,420**	1				
		p=,016	p<,001	p<,001	p<,001	p<,001					
	VE	-,068	,315**	,297**	,282**	,270**	,526**	1			
		p=,363	p<,001	p<,001	p<,001	p<,001	p<,001				
	BE	-,067	,216**	,246**	,078	,188**	,567**	,173*	1		
		p=,370	p=,002	p<,001	p=,269	p=,007	p<,001	p=,013			
	SY	-,121	,550**	,561**	,400**	,442**	,698**	,233**	,177*	1	
		p=,105	p<,001	p<,001	p<,001	p<,001	p<,001	p=,001	p=,011		
	SF	-,164	,188**	,148*	,269**	,135	,666**	,169*	,200**	,155*	1
		p=,028	p=,007	p=,035	p<,001	p=,054	p<,001	p=,015	p=,004	p=,026	

Alle Unterskalen korrelierten signifikant, bis auf „Beeinträchtigung“ (HoNOS) mit „zwischenmenschliche Beziehungen“ (EB-45) und „soziales Funktionsniveau“ (HoNOS) mit „soziale Integration“ (EB-45). Für GAF und HoNOS total score fand

sich auch wieder eine signifikante Korrelation, jedoch nicht mehr mit den Unterskalen.

3.4 Vergleich der Beurteilerperspektiven über die Zeit

3.4.1 Beeinträchtigungsschwere

In Tabelle 10 wurden die Mittelwerte der drei Instrumente gegenübergestellt. Es zeigte sich, dass der Mittelwert des GAF bei T_2 deutlich erhöht war im Vergleich zu T_1. Analog dazu waren im zeitlichen Verlauf die Mittelwerte von HoNOS total score und EB-45 TOT deutlich gesunken.

Tabelle 10: Gegenüberstellung der Mittelwerte (MW) und T-Test
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

	MW	Differenz	N
GAF			
T_1	44,16	t(188)= -17,339; p<,001	276
T_2	66,17		198
HoNOS total score			
T_1	11,73	t(206)=21,538; p<,001	294
T_2	5,60		207
EB-45 TOT			
T_1	79,03	t(217)=14,111; p<,001	294
T_2	56,81		218

Tabelle 11: T-Tests der Unterskalen von HoNOS und EB-45
Anmerkungen: EB-45 = Ergebnisfragebogen-45, HoNOS = Health of the Nation Outcome Scale.

		Differenz
HoNOS	Verhalten	t(206)=12,404; p<,001
	Beeinträchtigung	t(206)=09,004; p<,001
	Symptomatik	t(206)=17,541; p<,001
	Soziales Funktionsniveau	t(206)=10,346; p<,001
EB-45	Symptombelastung	t(217)=14,110; p<,001
	Zwischenmenschliche Beziehungen	t(217)= 9,058; p<,001
	Soziale Integration	t(217)=10,390; p<,001

Beim T-Test für gepaarte Stichproben ergab sich, dass alle Instrumente inklusive der jeweiligen Unterskalen im zeitlichen Verlauf eine signifikante Veränderung ($p<0,01$) erfahren hatten (siehe Tabelle 10 und Tabelle 11). Am deutlichsten ver-

änderte sich der HoNOS. Dies sieht man auch am Mittelwert, welcher sich zu T_2 halbierte.

3.4.2 Zusammenhänge zwischen den Perspektiven

In Abbildung 5 (S. 45) sind Streudiagramme zu den Korrelationen der Gesamtskalen der Instrumente mit der jeweiligen Regressionsgerade für T_1 und T_2 zu sehen. Die dazugehörigen Werte finden sich in den Tabellen 6 und 9. Es ist ersichtlich, dass die Korrelation im Verlauf von T_1 nach T_2 zwischen EB-45 TOT und HoNOS total score zunahm. Die Korrelation zwischen GAF und HoNOS total score nahm dagegen leicht ab. Wegen der negativen Korrelation, verursacht durch die gegenläufige Skalierung der beiden Instrumente, sinkt die Regressionsgerade hier. Wie bereits erwähnt, war die Korrelation von GAF und EB-45 TOT zu T_1 und T_2 nicht signifikant. Dies ist an den annähernd horizontal verlaufenden Regressionsgeraden in den beiden unteren Diagrammen zu sehen.

Zu T_2 korrelierte der GAF nicht mehr signifikant mit den Unterskalen der HoNOS („Beeinträchtigung“ und „soziales Funktionsniveau“). Die gute Korrelation der Unterskala „Symptomatik“ der HoNOS mit allen drei Unterskalen des EB-45 nahm bei T_2 zu.

In Tabelle 12 und Tabelle 13 (S. 46) sind die für das „Graphical Chain Modell“ berechneten angepassten Partialkorrelationen für die stationäre Aufnahme und die Entlassung angegeben. Eine geringe Korrelation zeigte sich zu T_1 für EB-45 TOT und HoNOS total score. Zu T_2 war diese Korrelation dann etwas deutlicher ausgeprägt. Ein Zusammenhang zwischen EB-45 TOT und GAF war zu beiden Zeitpunkten nicht gegeben. Ebenso war zwischen HoNOS total score und GAF nur zu T_1 eine schwache Beziehung zu finden.

Tabelle 12: Angepasste Partialkorrelationen bei Aufnahme (T_1)
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale.

T_1	**EB-45 TOT**	**HoNOS total score**	**GAF**
EB-45 TOT	1		
HoNOS total score	0,179	1	
GAF	0	-0,205	1

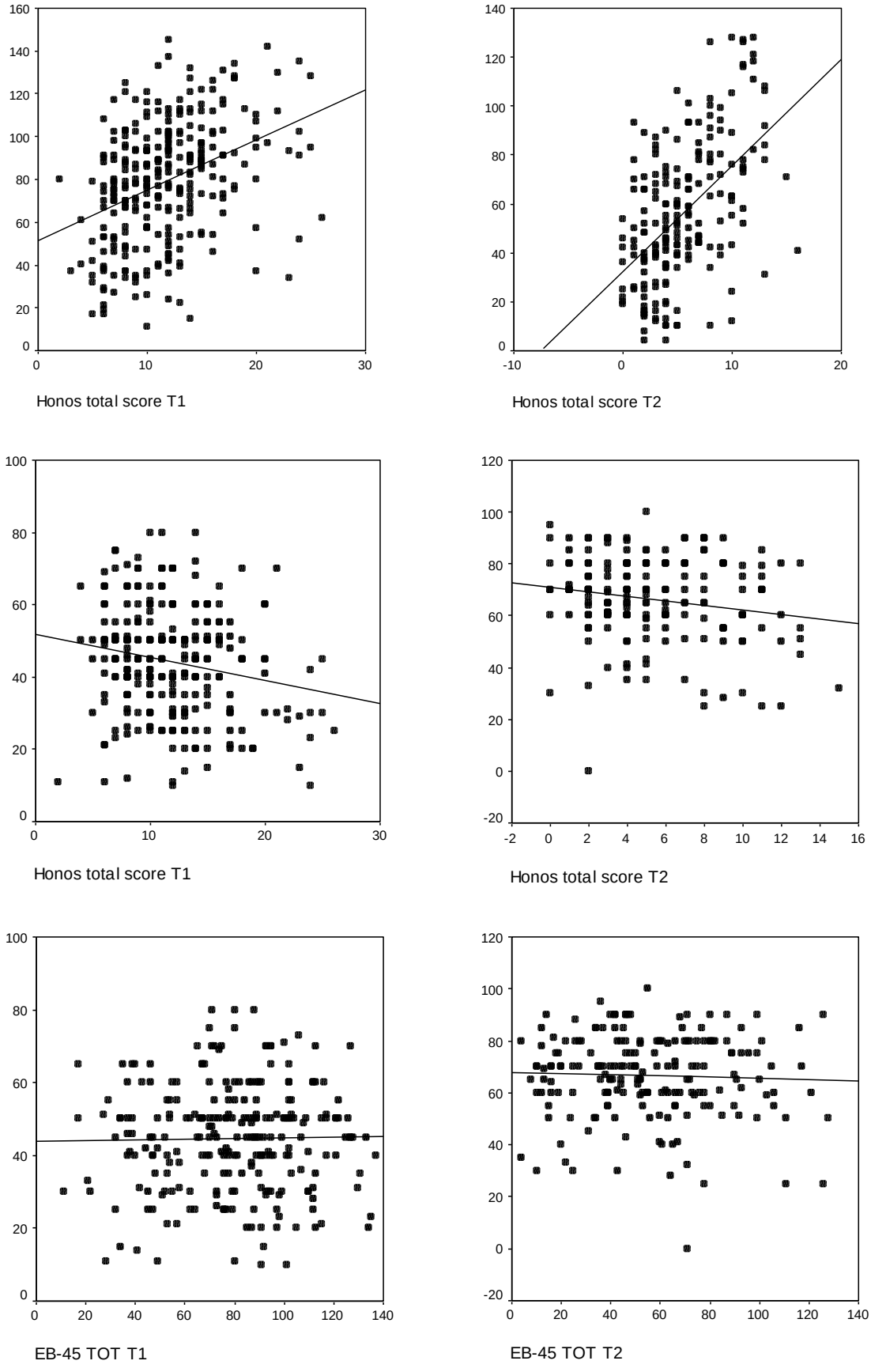

Abbildung 5: Streudiagramme zu den Korrelationen der Summenwerte
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T1 = Zeitpunkt der Aufnahme, T2 = Zeitpunkt der Entlassung.

Tabelle 13: Angepasste Partialkorrelationen bei Entlassung (T_2)
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale.

T_2	**EB-45 TOT**	**HoNOS total score**	**GAF**
EB-45 TOT	1		
HoNOS total score	0,421	1	
GAF	0	0	1

Tabelle 14 zeigt die angepassten Korrelationskoeffizienten für die Effekte über die Zeit von der Aufnahme bis zur Entlassung. Abgesehen von den guten Korrelationen zwischen den jeweiligen Aufnahme- und Entlassungswerten von GAF, HoNOS total score und EB-45 TOT, existierte nur ein schwacher und gegensinniger Zusammenhang zwischen EB-45 TOT zu T_1 und HoNOS total score zu T_2.

Tabelle 14: Angepasste Partialkorrelationen über die Zeit
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

	EB-45 TOT T_1	**HoNOS total score T_1**	**GAF T_1**
EB-45 TOT T_2	0,619	0	0
HoNOS total score T_2	-0,180	0,419	0
GAF T_2	0	0	0,400

In Abbildung 6 ist das zu den angepassten Partialkorrelationen dazugehörige „Graphical Chain Modell“ zu sehen. Wegen der schwachen Korrelation, gab es zwischen GAF und HoNOS total score keine Verbindung mehr zu T_2. Es existierte aber weiterhin eine direkte Verbindung zwischen EB-45 TOT und HoNOS total score, welche sich zu T_2 weiter verstärkte. Des Weiteren bestand weder zu T_1 noch zu T_2 eine direkte Verbindung zwischen EB-45 TOT und GAF.

Über die Zeit existierten Verbindungen zwischen den Eingangs- und Ausgangswerten von EB-45 TOT, HoNOS total score und GAF. Es gab nur eine Verbindung zwischen unterschiedlichen Instrumenten über die Zeit, und zwar zwischen EB-45 TOT und HoNOS total score, diese war jedoch gegensinnig und sehr schwach (-0,180).

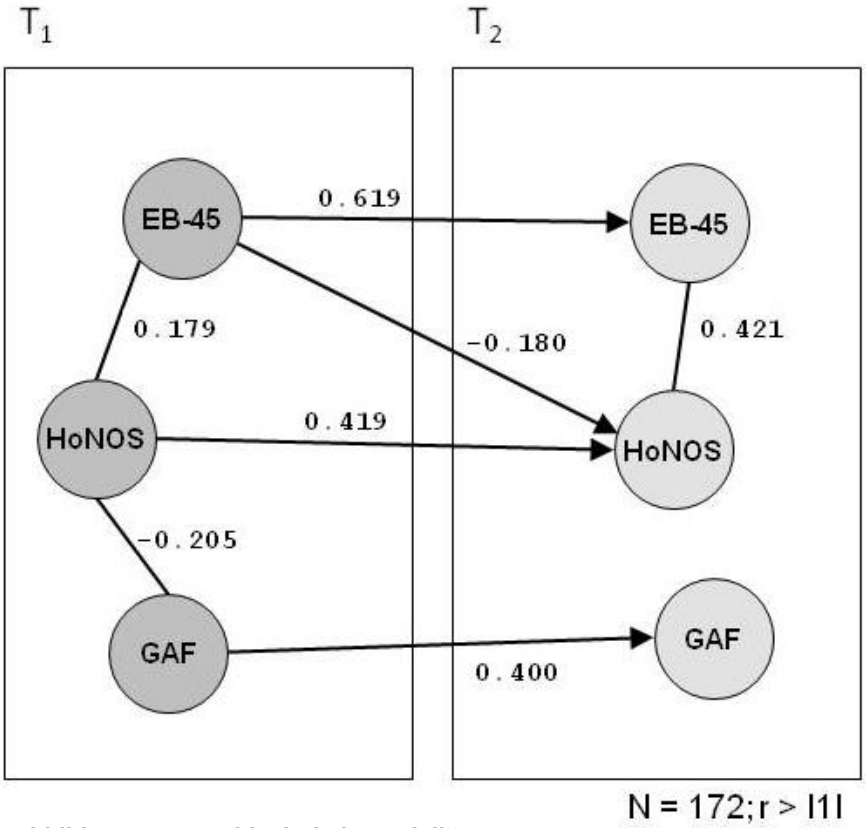

Abbildung 6: „Graphical Chain Modell“
Anmerkungen: EB-45 = Ergebnisfragebogen-45, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

3.4.3 Behandlungsergebnis

Zur besseren Veranschaulichung und Vergleichsmöglichkeit der relativen Häufigkeiten der als dysfunktional oder funktional eingestuften Teilnehmer über die Zeit dient die Abbildung 7 (S. 48). Hier sieht man, dass bei der Bewertung durch die Behandler (GAF) die größte Veränderung (5,6 %) zu verzeichnen war, gefolgt von den unabhängigen Bewertern (42,5 %). Die geringste Veränderung fand sich bei der Patienteneinschätzung (28,6 %). Wie am Chi^2-Test in Tabelle 15 zu sehen, war auf allen drei Skalen eine signifikante Veränderung der Anteile an dysfunktional und funktional bewerteten Teilnehmern zu verzeichnen.

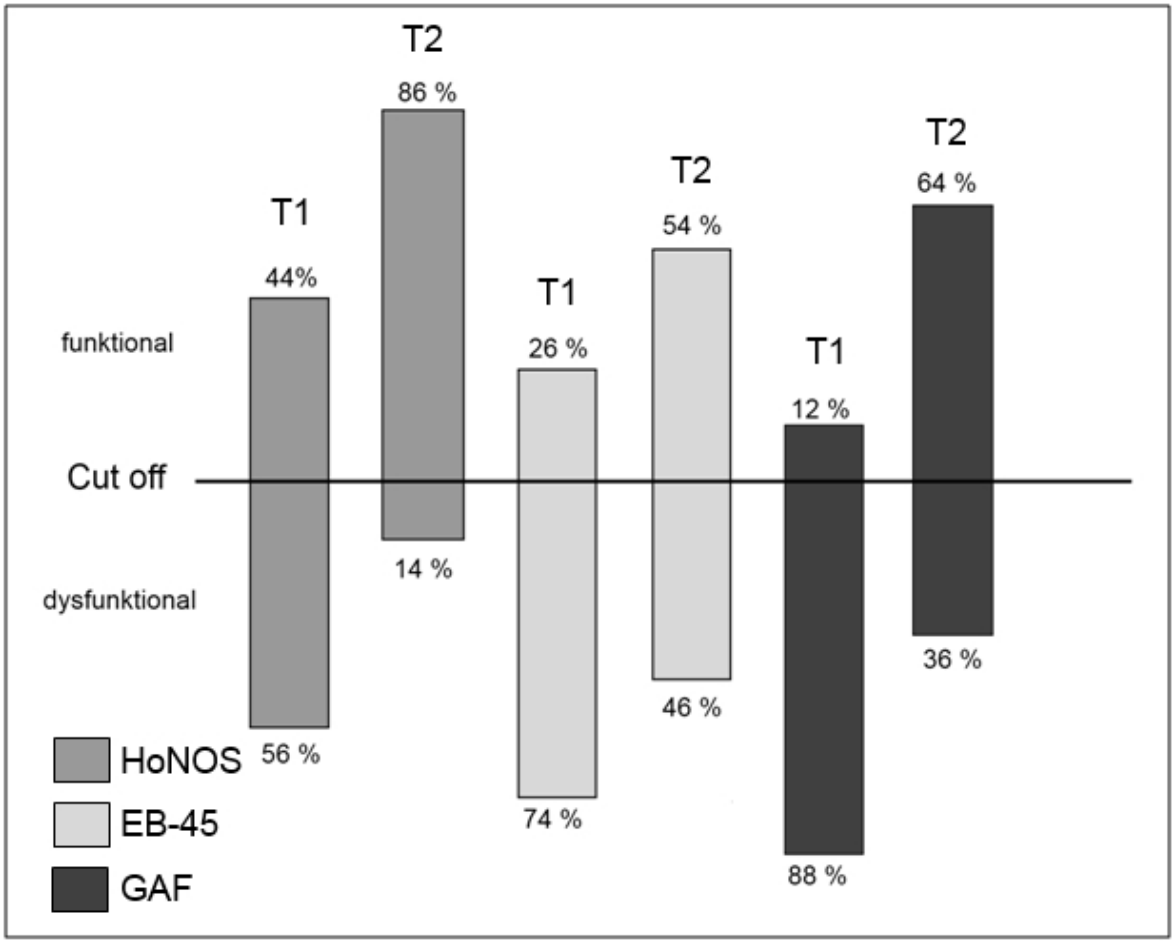

Abbildung 7: Relative Häufigkeiten der als dysfunktional und funktional eingestuften Teilnehmer für HoNOS total score, EB-45 TOT und GAF
Anmerkungen: EB-45 = Ergebnisfragebogen-45, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T1 = Zeitpunkt der Aufnahme, T2 = Zeitpunkt der Entlassung.

Tabelle 15: Chi-Quadrat-Test zur Einteilung in funktional und dysfunktional zu T_1 und T_2
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

	N	Differenz
HoNOS total score	214	$x^2_{(df=1)} = 11{,}212$; p=,001
EB-45 TOT	218	$x^2_{(df=1)} = 42{,}771$; p<,001
GAF	189	$x^2_{(df=1)} = 6{,}579$; p=,001

Was die übereinstimmende Einteilung der Teilnehmer in funktional und dysfunktional betrifft, lässt sich sagen, dass diese zwischen den Behandlern und den Patienten über die Zeit um acht Prozent abnahm (vgl. Abbildung 3, S. 38 und Abbildung 4, S. 41). Die übereinstimmenden Einschätzungen von Behandlern und Ex-

ternen bzw. Patienten und Externen blieben hingegen relativ konstant. Der höchste Grad an übereinstimmender Bewertung des psychischen Zustands fand sich zu T_1 zwischen Patienten und Behandlern.

In Tabelle 16 sind die absoluten und relativen Anteile der anhand des RCI Kriteriums [32] reliabel verbessert, reliabel verschlechtert und reliabel nicht verändert eingestuften Patienten für den EB-45 TOT, den GAF und den HoNOS total score angegeben. Der größte Anteil an reliabel verbesserten Patienten ergab sich bei der Auswertung der Behandlereinschätzung. Etwas mehr als die Hälfte aller Patienten schätzte sich selber als reliabel verbessert ein und der geringste Teil der Patienten wurde von den neutralen Beobachtern als reliabel verbessert eingestuft.

Tabelle 16: Reliable Veränderung und Gesundung
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale.

	N	Reliable Verbesserung N (%)	Reliable Verschlechterung N (%)	Keine reliable Veränderung N (%)	Gesundung N (%)
HoNOS total score	214	47 (22,0)	0 (0)	167 (78,0)	43 (20,1)
EB-45 TOT	218	113 (51,8)	6 (2,8)	99 (45,4)	63 (28,9)
GAF	189	134 (70,9)	2 (1,1)	53 (28,0)	90 (47,6)

Eine reliable Verschlechterung war bei allen drei Sichtweisen kaum oder gar nicht zu finden. Ebenso sind in Tabelle 16 die Anteile der Gesundeten aufgeführt. In den Abbildungen 8, 9 und 10 (S. 50/51) sind die dazu gehörigen Streudiagramme zu sehen (vgl. hierzu auch Abbildung 1, S.32). Die Punkte unterhalb des grauen Bandes repräsentieren in den Abbildungen 8 und 9 die reliabel Verbesserten und diejenigen darüber die reliabel Verschlechterten. In Abbildung 10 ist dies genau umgekehrt: Hier befinden sich die reliabel Verbesserten oberhalb des Bandes und die reliabel Verschlechterten unterhalb des Bandes. Wie in Tabelle 16 und Abbildung 9 ersichtlich, ergab sich für den HoNOS total score der geringste Anteil an gesundeten Patienten (N=43; 20,1 %). In der Abbildung 9 sind dies die Punkte im weißen Feld unten rechts. Für den EB-45 TOT wurden 63 „Gesundete“ (28,9 %) berechnet, welche in Abbildung 8 ebenfalls im weißen Feld unten rechts dargestellt sind. Weil der GAF entgegengesetzt skaliert ist, befinden sich die Punkte der „Gesundeten“ hier im weißen Feld oben links der Abbildung 10. Hier wurde mit 90 Patienten (47,6 %) die höchste Anzahl an „Gesundeten“ ermittelt.

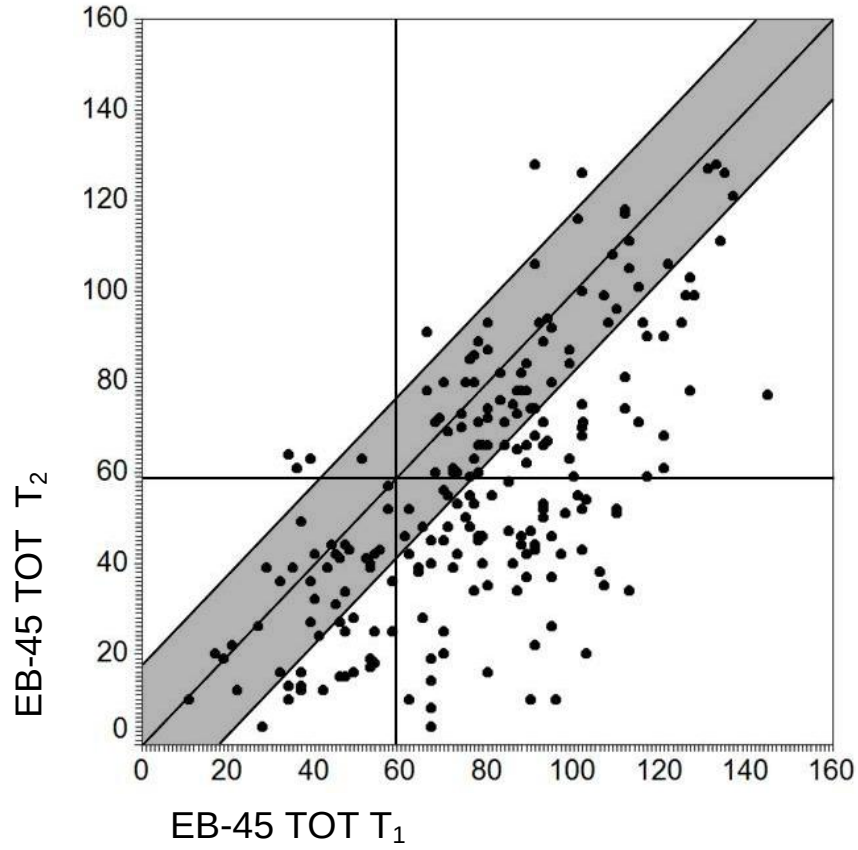

Abbildung 8: Scatterplot zur klinischen Signifikanz für EB-45 TOT
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

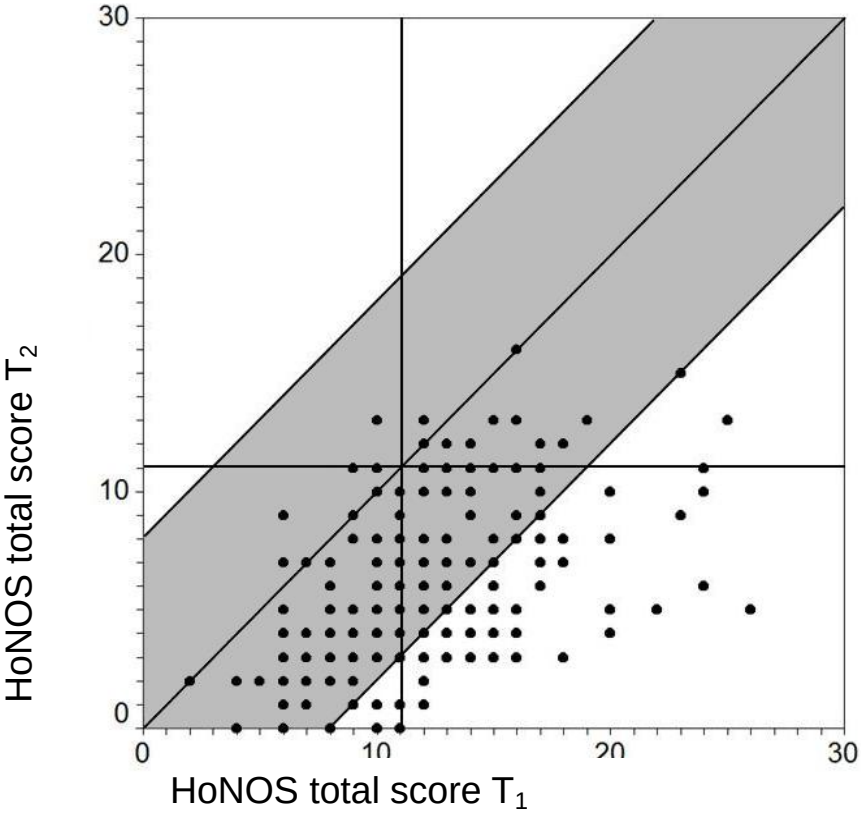

Abbildung 9: Scatterplot zur klinischen Signifikanz für HoNOS total score
Anmerkungen: HoNOS = Health of the Nation Outcome Scale, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

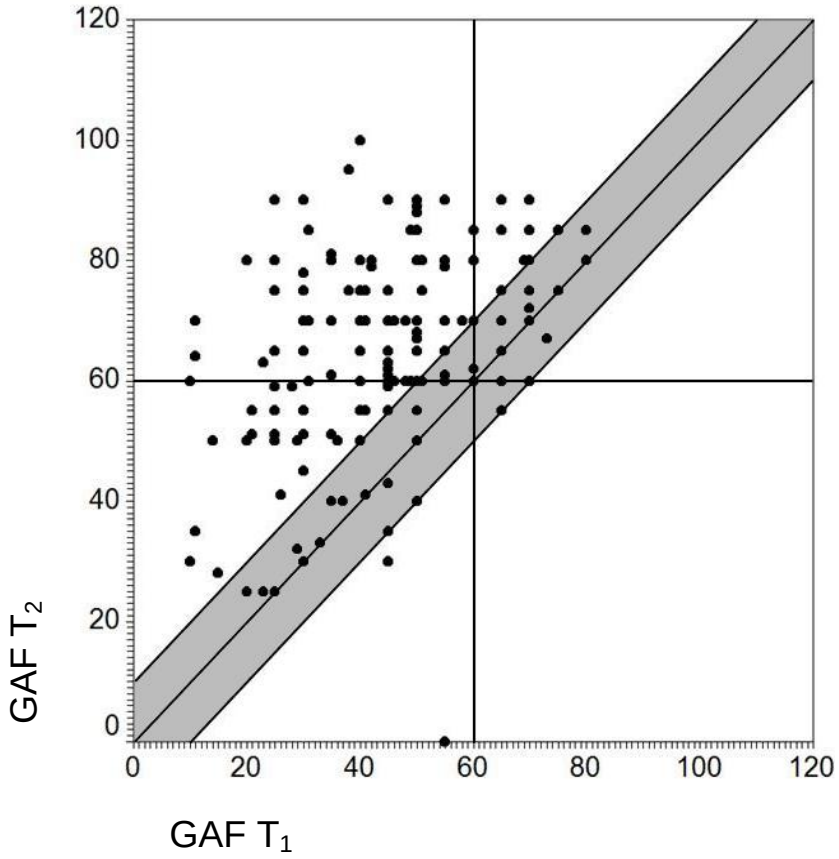

Abbildung 10: Scatterplot zur klinischen Signifikanz für GAF
Anmerkungen: GAF = Global Assessment of Functioning, T_1 = Zeitpunkt der Aufnahme, T_2 = Zeitpunkt der Entlassung.

Wie in Tabelle 17 zu sehen, war der Anteil an übereinstimmend als gesundet bewerteten Patienten für GAF und EB-45 TOT am höchsten. GAF und HoNOS total score besaßen schon deutlich weniger gleiche Bewertungen und für EB-45 TOT und HoNOS total score ergab sich die geringste Übereinstimmung an gesundeten Patienten.

Tabelle 17: Übereinstimmende Bewertungen an Gesundeten
Anmerkungen: EB-45 TOT = Ergebnisfragebogen-45 total score, GAF = Global Assessment of Functioning, HoNOS = Health of the Nation Outcome Scale.

	Gesamt N	**Gesundet N (%)**
GAF + HoNOS total score	180	20 (11,1)
EB-45 TOT + HoNOS total score	205	14 (12,0)
GAF + EB-45 TOT	182	32 (17,6)

4 Diskussion

4.1 Ergebnisdiskussion

4.1.1 Vergleich der Beurteilerperspektiven bei Aufnahme und Entlassung

Die erste Forschungsfrage dieser Arbeit war der querschnittliche Vergleich der Bewertungen von Behandlern, Patienten und neutralen Beobachtern zu Therapiebeginn und zur Entlassung stationär psychiatrisch-psychotherapeutisch behandelter Patienten. Zunächst wurde der psychische Gesundheitszustand der Patienten anhand von Cut-Offs in dysfunktional und funktional eingeteilt. Dies erfolgte für alle drei Skalen. Zum Zeitpunkt der stationären Aufnahme ergab sich der deutlich größte Anteil an funktional bewerteten Patienten für die HoNOS. Bei der Bewertung durch die Beobachter wurde fast die Hälfte der Teilnehmer als funktional eingestuft, obwohl sich diese erst zu Beginn ihrer Therapie befanden. Dem entgegengesetzt war ein sehr geringer Anteil an funktional eingeteilten Patienten zum Aufnahmezeitpunkt durch den GAF. Auffallend war weiterhin, dass die Selbstbeurteilung immer noch knapp die Hälfte an dysfunktionalen Patienten zu Therapieende ergab.

Der große Anteil an dysfunktional eingestuften Patienten bei stationärer Aufnahme aus Behandlersicht steht im Einklang mit dem Therapiebedarf, nicht jedoch die deutlich positivere Bewertung durch die externen Beurteiler. Nach ihrer Einschätzung müsste man davon ausgehen, dass bei knapp der Hälfte der Patienten gar keine Therapie nötig gewesen wäre. Der umgekehrte Fall war bei der Selbstbeurteilung eingetreten. Hier könnte man das Ergebnis so interpretieren, dass zum Zeitpunkt der Entlassung immer noch jeder zweite Patient dysfunktional, also erkrankt war, und deswegen einer weiteren Therapie bedurft hätte.

Zusammenfassend lässt sich sagen, dass die geschulten Rater vor und nach der Therapie positivere Bewertungen abgaben als die Behandler und die Patienten. Dies widerspricht der Studie von Ueki et al. [83], in der die Beobachter negativer bewerteten, als die Patienten mit nicht-endogener Depression und sich keine signifikanten Unterschiede in der Beurteilung der Beobachter mit Patienten mit endogener Depression ergaben. Hier waren die Beobachter jedoch keine geschulten

Rater, sondern Familienangehörige, wodurch diesen eine subjektivere Sichtweise zu unterstellen ist als den geschulten Ratern dieser Arbeit, deren Qualifikation und Objektivität Ursache für die unterschiedlichen Ergebnisse sein könnte [42]. Weitere Gründe dafür wären die verschiedenen Wissensstände der Familienangehörigen und geschulten Beurteiler dieser Arbeit, denn auch diese führen zu ungleichen Bewertungen [77].

Um zu untersuchen, inwieweit die Einordnungen der Patienten in funktional und dysfunktional zwischen den drei Perspektiven übereinstimmten, wurden Spezifitäten und Sensitivitäten berechnet. Die Übereinstimmungen waren bei stationärer Aufnahme für GAF und HoNOS total score und GAF und EB-45 TOT, sowie HoNOS total score und EB-45 TOT ungefähr gleich groß und mit ungefähr zwei Dritteln zu T_1 und T_2 auch relativ gut. Die größte Übereinstimmung bestand zwischen den Patienten und den Behandlern zu T_1 und die geringste zu diesem Zeitpunkt zwischen den Behandlern und den Beobachtern. Es fällt auf, dass nach erfolgter stationärer Therapie die Behandler etwas mehr als ein Viertel der Studienteilnehmer als funktional bewerteten, obwohl sich diese selber als dysfunktional einschätzten. Auch in der Studie von Trauer et al. [82], waren die Übereinstimmungen der einzelnen Items der HoNOS für Behandler und Patienten mäßig bis schlecht. Hingegen ergaben sich bei Ueki et al. [83] signifikante Übereinstimmungen zwischen Patienten und Beobachtern.

Anschließend wurden Korrelationen (Pearson) ermittelt. Es fanden sich zur Aufnahme und zur Entlassung gute signifikante Korrelationen zwischen HoNOS total score und EB-45 TOT, sowie der meisten ihrer Unterskalen. Die größte Korrelation der Unterskalen war zu T_2 zwischen „Symptomatik“ (HoNOS) und „Symptombelastung“ (EB-45). Die Korrelationen der Behandler mit den externen Bewertern waren signifikant (auch mit den Unterskalen „Beeinträchtigung“ und „soziales Funktionsniveau“) aber gering. Zwischen den Behandlern und den Patienten ergaben sich zu beiden Zeitpunkten keine signifikanten Korrelationen. Es bestand also anhand der Korrelationen ein Zusammenhang zwischen der Patientenbewertung und der Einschätzung der externen Beobachter und erstaunlicherweise kein signifikanter Zusammenhang zwischen den Behandlern und den Patienten. Dies stimmt mit der Aussage überein, dass Patienten und Behandler ihre Bedürfnisse unterschiedlich einschätzen [69] und auch mit derjenigen, dass Patienten und Behandler unter-

schiedliche Therapieziele setzen und deswegen zu unterschiedlichen Bewertungen gelangen [11, 39, 82]. Ein weiterer Grund für die Unterschiede könnte auch in der psychischen Erkrankung gelegen haben [18, 42, 71, 81, 88], denn die Patienten müssen bei der Selbstbeurteilung die Fähigkeit zur Selbstreflexion besitzen [77, 81, 84]. Stattdessen besitzen psychisch Kranke oft eine verzerrte Wahrnehmung und einen Mangel an Verständnis [15]. Bei Eichenberger et al. [15] waren die Korrelationen zwischen Behandler- und Patientenperspektive deutlich besser (r = 0,64 und r = 0,68 im Abstand von 6 Monaten) als in dieser Arbeit. Dies könnte daran liegen, dass dort die Studienteilnehmer aus ambulanter Betreuung stammten und demzufolge eine geringere Psychopathologie aufwiesen. Dem widerspricht jedoch die Untersuchung von Schneider et al. [71] in der, ebenso wie bei den Ergebnissen dieser Arbeit, kein korrelativer Zusammenhang zwischen der Arztbewertung und der Einschätzung ambulant betreuter Patienten vorlag.

4.1.2 Vergleich der Beurteilerperspektiven über die Zeit

Die zweite Forschungsfrage, die behandelt wurde, war der longitudinale Vergleich der drei Beurteilerperspektiven. Hierbei kamen zunächst Methoden zum Einsatz, die eine Aussage über die statistische Signifikanz ermöglichen. Anschließend wurden die Ergebnisse auf klinische Signifikanz untersucht.

Vergleich der Beeinträchtigungsschwere

In dieser Studie erfassten alle drei Messinstrumente eine deutliche Veränderung der Beeinträchtigungsschwere vor und nach der Therapie, weil eine signifikante Veränderung der Mittelwerte von HoNOS total score, EB-45 TOT (und deren Unterskalen) und GAF zwischen T_1 und T_2 stattfand. Die Veränderungen der Mittelwerte lassen auf eine Reduzierung der Beeinträchtigungsschwere nach erfolgter Therapie schließen. Folglich stimmten die Sichtweisen von Beobachtern, Patienten und Behandlern hierin überein. Wie bereits in anderen Studien festgestellt, bestätigt dies die gute Veränderungssensitivität von HoNOS, GAF und EB-45 [4, 29, 51]. Dem widersprechend war die Veränderung der Mittelwerte des GAF bei Ruggeri et al. [66] zwischen den beiden Erhebungszeitpunkten nicht signifikant. Es ergab sich dort keine Verbesserung der Psychopathologie im Laufe der Zeit. Dies ist aber wohl damit zu erklären, dass die zweite Erhebung mit deutlichem Abstand zur Therapie, nämlich zur 3-Jahres-Katamnese, stattfand.

Vergleich der Zusammenhänge zwischen den Perspektiven

Es zeigte sich bei Betrachtung der bereits weiter oben erwähnten Korrelationen (Pearson), dass diese zwischen der Selbstbeurteilung und der Fremdbeurteilung durch die Externen im Verlauf von der stationären Aufnahme bis zur Entlassung zunahm. Hier nahmen auch die signifikanten Korrelationen der HoNOS mit den Unterskalen des EB-45 zu. Dagegen nahm der bereits schwache Zusammenhang zwischen den Behandlern und den geschulten Ratern weiter ab. Bei diesen gab es auch keine Korrelationen mehr mit den Unterskalen der HoNOS, „Beeinträchtigung“ und „soziales Funktionsniveau“. Für die Behandler und die Patienten bestanden zu beiden Erhebungszeitpunkten keine signifikanten Korrelationen. Bezogen auf die angepassten Partialkorrelationen und das „Graphical Chain Modell“ änderte sich der eben geschilderte Zustand dahingehend, dass hier zu Therapieende nun keine signifikante Korrelation mehr zwischen den Behandlern und den geschulten Ratern vorhanden war.

Die Korrelationsanalyse zeigte, dass die Patienten- und Beobachterbewertungen miteinander enger verknüpft waren als die Behandlersichtweise mit der Sichtweise der Beobachter, die wiederum einander näher waren als die Behandler- und Patientenperspektive. Die Behandlereinschätzung war sogar deutlich isoliert von denen der Patienten und den geschulten Rater und ihre Verknüpfung zu den anderen Bewertungsskalen nahm über die Zeit eher noch ab. Die Verbindung der Patienten mit den geschulten Ratern hingegen war nach erfolgter stationärer Behandlung größer geworden. Dies lässt vermuten, dass hier die Verbesserung der Symptomatik einen Einfluss auf die Urteilsfähigkeit der Patienten hatte. So wurde schon mehrmals postuliert, dass die Diagnose und die Schwere der Erkrankung einen Einfluss auf die Beurteilung haben [16, 42, 71, 77, 81, 88]. Allerdings fanden sich hierzu auch widersprüchliche Ergebnisse. In zwei Untersuchungen hatte die Schwere der Erkrankungen bei Schizophrenie-Patienten keinen Einfluss auf das Ausfüllverhalten [82, 84] und in einer anderen Studie zeigte sich, dass sich Wahrnehmungsstörungen nachteilig auf das Ausfüllverhalten auswirkten [71].

Behandlungsergebnis

Für die Anteile an dysfunktional und funktional bewerteten Teilnehmern ergab sich eine signifikante Veränderung über die Zeit. Wie zu erwarten war, nahm der Anteil an funktional bewerteten Patienten zu Therapieende für alle drei Skalen deutlich

zu. Auffallend war, dass die Beurteilung durch die Behandler stark polarisiert war. Deswegen sah man bei Betrachtung der Differenzen der als dysfunktional eingeteilten Teilnehmer zwischen T_1 und T_2, dass die größte Verbesserung bei der Behandlersichtweise zu verzeichnen war, gefolgt von den externen Beobachtern, während sich bei der Patientenperspektive nur eine geringe Änderung ergab, so dass hier die geringste Verbesserung bestand. Auch laut Lambert et al. [42] wurde bei der Patientensichtweise weniger Veränderung der Symptomatik gemessen als bei der Behandlersichtweise. Dort wurde dies damit begründet, dass depressive Patienten ihre Symptome nicht korrekt einschätzen können [61]. Ein wichtiger Aspekt wird hier auf jeden Fall gewesen sein, dass die Behandler wussten, dass die Patienten eine Therapie erfahren hatten, und deswegen bei Therapieende besonders positiv bewerteten. Diesen Einfluss könnte man auch bei den Behandlern der vorliegenden Untersuchung vermuten. In der Studie von Schneider et al. [71] wurde ebenfalls eine deutliche Verbesserung der Symptomatik durch die Ärzte erfasst und bei deren Patienten nur ungefähr halb so viel.
Bei der Einteilung in funktional und dysfunktional war auffallend, dass die übereinstimmenden Behandler- und Patientenbewertungen im zeitlichen Verlauf um acht Prozent abnahmen. Hier hätte man eher erwartet, dass diese weiter zunehmen, wenn man, wie weiter oben bereits erwähnt, davon ausgeht, dass die Patienten nach erfolgter Therapie durch die reduzierte Symptomschwere bessere Selbstbeurteilungen abgeben können. Entweder wichen die Bewertungen voneinander ab, weil die Behandler, wie auch schon erwähnt, unbewusst einen Therapieeffekt sehen wollten oder weil die Patienten trotz Therapie aufgrund ihrer Erkrankung immer noch keine richtige Selbsteinschätzung durchführen konnten. Eine weitere Begründung für die unterschiedlichen Bewertungen von Patienten und Behandlern könnte darin liegen, dass diese die Bedürfnisse des Patienten und damit auch Ansprüche an das Behandlungsergebnis unterschiedlich einschätzten [69]. Denn Stimmungen und individuelle Bedürfnisse des Patienten sind für Fremdbeurteiler nicht zu bewerten [74, 75, 77]. Vielmehr konzentrieren sich die Behandler auf die Verbesserung der Symptomatik [82].

Die zentrale Auswertungsmethode dieser Untersuchung war die Methode der klinischen Signifikanz nach Jacobson et al. [33]. Hierbei wurde das Behandlungsergebnis in folgende Bereiche geteilt: Gesundung, keine Veränderung, reliable Ver-

besserung und reliable Verschlechterung. Die deutlich größte reliable Verbesserung (nach RCI) mit über zwei Dritteln war für den GAF zu verzeichnen, gefolgt vom EB-45 TOT mit mehr als die Hälfte reliabel verbesserter Patienten. Für den HoNOS total score zeigte sich nur eine sehr geringe Zahl an verbesserten Teilnehmern. Die größte Zahl an gesundeten Patienten errechnete sich für die Behandlerperspektive. Fast jeder zweite Patient galt hier am Ende als gesundet. Bei der Selbstbeurteilung zeigte sich jeder Vierte als gesundet und bei den Außenstehenden nur jeder fünfte Teilnehmer. In der Literatur lag ein gemischtes Bild vor, was die Bewertungen des Behandlungsergebnisses verschiedener Sichtweisen betraf. Die vorliegende Untersuchung stimmt mit den Studien von Lambert et al. [42], Schneider et al. [71] und Trauer et al. [82] überein, in denen die Patienten das Behandlungsergebnis negativer einschätzten als die Behandler. Der vorliegenden Arbeit widersprechend ist die Studie von Eichenberger et al. [15], welche Patienten mit chronischer Schizophrenie untersuchten und feststellten, dass hier die Behandler im Vergleich zu ihren Patienten insgesamt negativere Bewertungen abgaben. Auch bei Ruggeri et al. [66] bewerteten die Behandler das Behandlungsergebnis häufiger schlechter als die Patienten. Für den GAF ergaben sich hier 47 % verschlechterte, 23 % stabile und 30 % verbesserte Patienten. Bei den Patienten wurde das Behandlungsergebnis über die Lebensqualität bestimmt (LQoLP). Für 24 % ergab sich eine verschlechterte, für 57 % eine stabile und für 19 % eine verbesserte Situation. Der größte Wert an übereinstimmend als gesundet bewerteten Teilnehmern zeigte sich in der vorliegenden Untersuchung für GAF und EB-45 TOT und der geringste für EB-45 TOT und HoNOS total score. Insgesamt waren die Übereinstimmungen als eher gering anzusehen. Bei Voruganti et al. [84] ergaben sich dagegen hohe Übereinstimmungen zwischen den Patienten und dem klinischen Fachpersonal. Diese waren besonders groß, wenn der Behandler mit der Lebenssituation des Patienten vertraut war. Ein deutlicher Unterschied zur vorliegenden Arbeit war, dass die Patienten- und Behandlerperspektiven mithilfe von Interviews durch einen geschulten Studienmitarbeiter erfasst wurden. Dadurch war die Motivation der Patienten verbessert [81], wodurch sich wiederum eine Tendenz zu einer etwas positiveren Selbsteinschätzung ergeben haben könnte [83].

Im Vergleich zu den vorhandenen Studien zeigte sich also, dass die zentralen Arbeiten der Literatur zu teilweise anderen Ergebnissen gelangten, wobei diese auch

untereinander verschieden waren. Oft wurden nur einzelne Diagnosegruppen (v.a. Depressionen oder Schizophrenie) untersucht und auch die Stichprobenumfänge fielen unterschiedlich (teilweise sehr klein, z.B. N = 37) aus. Ebenso kamen unterschiedliche Messinstrumente zum Einsatz, die in ihren inhaltlichen Schwerpunkten und ihrer psychometrischen Qualität verschieden waren. Darüber hinaus erfolgte die Auswertung der Daten mit teils unterschiedlichen statistischen Verfahren.

4.2 Methodendiskussion

Diese Untersuchung verwendete verschiedene statistische Verfahren, die sich in der Art der untersuchten Signifikanz unterscheiden. Der T-Test, die Korrelationsanalyse und das „Graphical Modelling“ sind Methoden, die die Ergebnisse auf statistische Signifikanz untersuchen. Die Einteilung des psychischen Zustands in dysfunktional und funktional, sowie die Spezifität und Sensitivität berücksichtigen die klinische Bedeutung der Messwerte während die Methode der klinischen Signifikanz nach Jacobson et al. [33] diese mit der reliablen Veränderung kombiniert. Es fanden sich über die Korrelationsanalyse und das „Graphical Modelling“ die höchsten Übereinstimmungen zwischen den Patienten und den neutralen Beobachtern und über die Spezifität/Sensitivität und klinische Signifikanz die höchsten Übereinstimmungen zwischen Patienten und Behandlern. Unterschiedliche Ergebnisse für statistische und klinische Signifikanz wurden auch in einer Studie von Moleiro et al. [50] erfasst. Die unterschiedlichen Ergebnisse sind damit zu erklären, dass die statistische Signifikanz nichts über die Bedeutung des Ergebnisses für den Patienten aussagt [50, 59]. Denn der p-Wert macht keine Aussage über die Richtung oder Größe eines Ergebnisses [18, 59] und ebenso wenig über die Ursachen und Konsequenzen einer Signifikanz [68]. Ein weiterer Kritikpunkt des p-Wertes besteht darin, dass er von der Fallzahl abhängig ist. Das bedeutet, dass auch kleine Differenzen, die klinisch nicht signifikant sind, bei ausreichend großen Stichproben durchaus statistisch signifikant sein können [18]. Außerdem kann eine zu große Streuung der Daten (zum Beispiel durch eine heterogene Patientengruppe) zu nicht statistisch signifikanten Ergebnissen führen [59]. Ein zentraler Unterschied zwischen T-Test und klinischer Signifikanz ist, dass dieser nur eine Aussage über die Gesamtstichprobe macht und nicht über den einzelnen Patienten. Das Ergebnis der Korrelation macht gar keine Aussage über das Behandlungsergebnis. Sie

sagt nichts über die tatsächlichen Zusammenhänge der Instrumente aus, denn der Korrelationskoeffizient drückt nur aus, dass ein Zusammenhang nicht auszuschließen ist. Des Weiteren gilt es zu beachten, dass auch geringe Korrelationen eine große klinische Bedeutung besitzen können, besonders wenn es um dichotome Ergebnisse geht (z.B. gesundet oder nicht gesundet) [68].

4.3 Limitationen

Ein Nachteil dieser Untersuchung war, dass die Vergleichbarkeit der Instrumente eingeschränkt ist. Alle Instrumente waren sowohl vom Umfang als auch inhaltlich verschieden. Dadurch dürfte auch ihre inhaltliche Aussagekraft unterschiedlich sein. Das wurde auch schon von Salvi et al. [70] gezeigt. Diese verglichen vier Instrumente aus Behandlersicht miteinander, die im Ergebnis keine grundlegenden Übereinstimmungen untereinander besaßen.
Die Verschiedenheit der Instrumente hatte auch praktische Konsequenzen für die statistische Auswertung. So konnte z. B. kein Vergleich der Mittelwerte durchgeführt werden, weil die Summenwerte der verwendeten Skalen unterschiedliche Bedeutungen haben. Des Weiteren könnte denkbar sein, dass durch die Unterschiedlichkeit der Instrumente selber so große Schwankungen erzeugt wurden, dass diese größer waren als der eigentliche Therapieeffekt [42].
Außerdem ist es eine schwierige Frage, ab welchem Wert ein Patient als erkrankt oder gesund einzustufen ist [33, 50]. Für diese Untersuchung wurden die Cut-Offs für HoNOS und GAF aus der Literatur übernommen und für EB-45 selbst ermittelt. Als Referenz wurde dabei eine Bevölkerungsstichprobe aus Deutschland verwendet. Je nachdem wie die Ergebnisse solch einer Stichprobe ausfallen, so verändert sich auch der Cut-Off. Deswegen fanden sich in der Literatur zu einzelnen Instrumenten auch unterschiedliche Vorschläge für Cut-Offs. Je nach Wahl des Cut-Offs verschiebt sich jedoch das Verhältnis von funktional zu dysfunktional [45]. Das Gleiche trifft auch für die Festlegung der kritischen Differenzen zu. Den Cut-Off richtig festzulegen ist oft schwierig, weil meistens keine Normstichproben zum Vergleich vorliegen. Man kann deswegen häufig nicht beurteilen, ob die Stichprobe einer Studie repräsentativ ist oder nicht [32]. Bei der Festlegung einer Normalpopulation ist auch nie auszuschließen, dass sich darunter auch unentdeckte „Dysfunktionale“ befinden [91]. Weiterhin gilt es bei der Ermittlung des Cut-Offs zu

bedenken, dass es im psychiatrischen Bereich oft unrealistisch ist einen Patienten so zu therapieren, dass er die gleichen Werte wie die der Normalpopulation erhält. Es ist oft schon ein klinischer Erfolg, wenn der Patient einen Zustand erlangt, in dem er wieder entlassungsfähig ist [32].
Außerdem ist der GAF aufgrund seiner Kürze für diese Art der Untersuchung keine gut geeignete Skala. Auch seine Reliabilität ist in Frage zu stellen, denn im Vorfeld hatte kein Training zu seiner Nutzung stattgefunden.

4.4 Stärken der Studie

Die Studie zeichnet sich durch die Erhebung aus drei verschiedenen Perspektiven (Patienten, Behandler und neutrale Beobachter) aus. In den meisten ähnlichen Studien wurden nur zwei Sichtweisen verglichen.
Eine weitere Stärke stellt die Stichprobengröße von 294 Probanden dar, die zum zweiten Erhebungszeitpunkt immerhin noch aus 226 teilnehmenden Personen bestand.

4.5 Schlussfolgerung und Ausblick

Insgesamt lässt sich sagen, dass die Bewertungen der drei Perspektiven eher unabhängig voneinander waren und sich unterschiedliche Behandlungsergebnisse für die drei Perspektiven ergaben. Damit stehen diese Ergebnisse im Einklang mit denen von Krieble et al. [39] und Stieglitz et al. [77], welche beide postulierten, dass aufgrund der unterschiedlichen Blickwinkel der verschiedenen Bewerter nur eine begrenzte Korrelation vorliegen kann. Keine der drei Sichtweisen ist also als falsch oder richtig bezogen auf die Anderen anzusehen. Um das Behandlungsergebnis zu erfassen, kann deswegen empfohlen werden stets mehrere Perspektiven zu erfassen. Anhand der Ergebnisse lässt sich weiter schließen, dass die Patienten fähig sind an Ergebnismessung teilzunehmen. Ziel der Versorgungsforschung ist es u. a. die Qualität und Effizienz der Therapien zu verbessern. Deswegen sollten die Patienten auch vermehrt in die Behandlungsplanung eingebunden werden.
Die hier v.a. beim „Graphical Chain Model“ ermittelte isolierte Stellung der Behandlereinschätzung gibt einem zu denken. Die Ursachen hierfür, wie z. B. eine zu

starke Konzentration auf die Symptombelastung der Patienten, bedürfen weiterer Forschung. In zukünftigen Arbeiten wäre es außerdem sinnvoll die Faktoren (wie z.B. Alter, Diagnose, Compliance, Stimmung) zu untersuchen bzw. zu ermitteln, die einen Einfluss auf die Bewertung des Behandlungsergebnisses haben könnten. Dabei wäre es empfehlenswert einheitliche Fragebögen zu verwenden bzw. solche, die in Versionen für verschiedene Perspektiven vorliegen (z.B. HoNOS). Auch der Einfluss der verwendeten statistischen Methoden, insbesondere die unterschiedlichen Ergebnisse für die statistische und klinische Signifikanz, sollten weiter untersucht werden und Studienergebnisse im Hinblick darauf kritisch betrachtet werden. Von der strengen Konzentration auf statistisch signifikante Ergebnisse sollte ein Umdenken in Richtung praktischer und realer Bedeutung der Messwerte stattfinden.

5 Zusammenfassung

Die in den letzten Jahrzehnten entstandene Versorgungsforschung hat es sich zum Ziel gesetzt die psychiatrisch-psychotherapeutische Versorgung zu verbessern. Um dies zu erreichen, wurden und werden Methoden entwickelt, die das Behandlungsergebnis bzw. den aktuellen Behandlungsstatus mit Hilfe von Fragebögen erfassen.

In der vorliegenden Arbeit wurden das Behandlungsergebnis und die Krankheitsschwere von 294 stationär und psychiatrisch-psychotherapeutisch behandelten Patienten aus drei Beurteilerperspektiven verglichen. Diese drei Perspektiven waren Patienten, Behandler und geschulte Rater. Zur Aufnahme und zur Entlassung wurden diese drei Sichtweisen auf Unterschiede, Gemeinsamkeiten und Zusammenhänge untersucht. Anschließend fand ein Vergleich dieser Bewertungen im zeitlichen Verlauf statt und die jeweiligen Behandlungsergebnisse wurden ermittelt.

Die Probanden wurden an der „Klinik für Psychiatrie, Psychotherapie und Psychosomatik“ des Bezirkskrankenhauses Günzburg im Rahmen einer cluster-randomisierten Studie („Ergebnismonitoring und Ergebnismanagement in der stationären psychiatrischen Versorgung“ = EMM-Studie) rekrutiert. Die Daten zur Krankheitsschwere wurden anhand von Fragebögen erfasst, welche von den Behandlern, geschulten Mitarbeitern der Studie und den Patienten ausgefüllt wurden.

Neben deskriptiven statistischen Methoden, wurden die Daten mit Verfahren der statistischen Signifikanz (T-Test, Korrelationen und „Graphical Chain Modell“) und der Methode der klinischen Signifikanz untersucht.

Je nach Untersuchungsmethode ergaben sich teilweise unterschiedliche Ergebnisse. Für die statistische Signifikanz ergab sich, dass die Patienten- und die Beobachterbewertungen signifikant korrelierten. Ihre Verknüpfung nahm im zeitlichen Verlauf sogar zu. Zwischen den Patienten und ihren Behandlern bestand zu beiden Zeitpunkten und im zeitlichen Verlauf keine Verbindung im „Graphical Chain Modell“. Bei der Untersuchung auf klinische Signifikanz zeigte sich der größte Anteil an Patienten, die sich von einem dysfunktionalen zu einem funktionalen psychischen Gesundheitszustand verbesserten, bei der Behandlerperspektive und der geringste bei den Patienten selber. Die meisten Gesundeten ergaben sich bei den Behandlern und die wenigsten bei den geschulten Ratern. Der Anteil an über-

einstimmend als gesundet bewerteten Patienten war für die Patienten und Behandler am größten und für die Patienten und geschulten Rater am geringsten. Allerdings bewegten sich diese Werte mit 11,1 -17,6 % in einem niedrigen Bereich. Auch insgesamt lässt sich festhalten, dass die Korrelationen und Übereinstimmungen zwischen den Perspektiven gering waren.

Aufgrund der unterschiedlichen Blickwinkel der verschiedenen Bewerter kann nur eine begrenzte Übereinstimmung der Ergebnisse vorliegen. Deswegen wird empfohlen zur Bestimmung des Behandlungsergebnisses mehrere Perspektiven zu erfassen. Auch wegen der teilweise isolierten Stellung der Behandlersicht, sollten in Zukunft nicht nur Bewertungen durch die Behandler alleine stattfinden, sondern immer auch durch die Patienten, denn diese scheinen trotz ihrer psychischen Beeinträchtigung durchaus dazu fähig und können damit stärker in ihre Behandlung und deren Verlaufsplanung eingebunden werden. Dies würde auch den zentralen Zielen der Versorgungsforschung, die Qualität und die Effizienz der Versorgung zu verbessern, zugutekommen.

Da für jede Perspektive ein anderes Instrument verwendet wurde, war deren Vergleichbarkeit eingeschränkt. Deswegen wäre es in zukünftigen Untersuchungen sinnvoll Parallelversionen von Fragebögen zu benutzen. Des Weiteren wäre es interessant die Faktoren zu untersuchen, von denen das Ausfüllverhalten beeinflusst wird. Die unterschiedlichen Ergebnisse für statistische und klinische Signifikanz sprechen für eine kritische Betrachtung von statistisch signifikanten Ergebnissen im Hinblick auf ihre klinische Bedeutung.

6 Literaturverzeichnis

1. Amador X, Strauss D H, Yale S H, Flaum M M, Endicott J, Gorman JM : Assessment of insight in psychosis. Am J Psychiatry 150: 873-880 (1993)

2. American Psychiatric Association: Diagnostic and Statistical Manual of Mental disorders. DSM-IV-TR. 4.Auflage, American Psychiatric Association, Washington DC, (2000)

3. Andreas S, Harfst T, Dirmaier J, Kawski S, Koch U, Schulz H: Verband Deutscher Rentenversicherungsträger (VDR) Landesversicherungsanstalt Rheinprovinz (Hrsg) In: Die „Health of the Nation Outcome Scale“: Entwicklung und erste psychometrische Analysen einer deutschen Version (HoNOS-D). DRV-Schriften 52: 31-33 (2004)

4. Andreas S, Harfst T, Rabung S, Mestel R, Schauenburg H, Hausberg M, Kawski S, Koch U, Schulz H: The validity of the German version of the Health of the Nation Outcome Scales (HoNOS-D): a clinician-rating for the differential assessment of the severity of mental disorders. Int J Methods Psychiatr Res 19: 50-62 (2010)

5. Beck A T, Ward C H, Mendelsohn M, Mock J, Erbaugh J: An inventory for measuring depression. Arch Gen Psychiatry 4: 561-571 (1961)

6. Bergner M, Bobbit R A, Carter W B, Gilson B S: The sickness impact profile: development and final revision of a health status measure. Med Care 19: 787-805 (1981)

7. Berner P, Gabriel E, Katschnig H, Kieffer W, Koehler K, Lenz G, Simhandl C (Hrsg) Diagnosekriterien für schizophrene und affektive Psychosen. Weltverband für Psychiatrie, Wien, (1983)

8. Brooks R: EuroQol: The current state of play. Health Policy 37: 53-72 (1996)

9. Buckingham W, Burgess P, Solomon S, Pirkis J, Eagar K: Casemix Classification of Mental Health Services. Commonwealth Department of Health and Family Services, Canberra, (1998)

10. Callaly T, Trauer T, Hantz P: Integration of outcome measures into clinical practice. Australas Psychiatry 6: 188-190 (1998)

11. Connolly M B, Strupp H H: Cluster analysis of patient reported psychotherapy outcomes. Psychother Res 6: 30-42 (1996)

12. Cording C, Gaebel W, Spengler A, Stieglitz R D, Geiselhart H, John U, Netzold D W, Schönell H: Die neue psychiatrische Basis-Dokumentation. Eine Empfehlung der DGPPN zur Qualitätssicherung im (teil-)stationären Bereich. Spektrum der Psychiatrie und Nervenheilkunde 24: 3-41 (1995)

13. Derogatis L R, Fasth B G: Confirmation of the dimensional structure of the SCL-90: a study in construct validation. J Clin Psychol 33: 981-989 (1977)

14. Edwards D: Introduction to Graphical Modelling. Springer, New York, (2000)

15. Eichenberger A, Rössler W: Comparison of self-ratings and therapist ratings of outpatients' psychosocial status. J Nerv Ment Dis 188: 297-300 (2000)

16. Eisen S V, Dickey B: Mental health outcome assessment: The new agenda. Psychotherapy (Chic) 33: 181-189 (1996)

17. Eisen S V, Dill D L, Grob M C: Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. Hosp Community Psychiatry 45: 242-247 (1994)

18. Eisen S V, Ranganathan G, Seal P, Spiro A: Measuring clinically meaningful change following mental health treatment. J Behav Health Serv Res 34: 272-289 (2007)

19. Endicott J, Spitzer R L, Fleiss J L, Cohen J: The Global Assessment Scale. Arch Gen Psychiatry 33: 766-771 (1976)

20. Gaebel W, Janssen B, Zielasek J: Mental health quality, outcome measurement, and improvement in Germany. Curr Opin Psychiatry 22: 636-642 (2009)

21. Goldberg D P: Manual of the General Health Questionnaire. Slough: NFER/Nelson, (1978)

22. Gurin G, Verhoff J, Feld S: Americans view their mental health. Russell Sage Foundation, New York (NY), (1960)

23. Hamilton M: A rating scale for depression. J Neurol Neurosurg Psychiatry 23: 56-62 (1960)

24. Hansson L: Outcome assessment in psychatric service evaluation. Soc Psychiatry Psychiatr Epidemiol 36: 244-248 (2001)

25. Hansson L, Björkmann T, Priebe S: Are important patient-rated outcomes in community mental health care explained by only one factor? Acta Psychiatr Scand 116: 113-118 (2007)

26. Hansson L, Höglund E: Patient satisfaction with psychiatric services. The development, reliability and validity of two patient satisfaction questionnaires for use in inpatient and outpatient settings. Nord J Psychiatry 49: 257-262 (1995)

27. Hasler G, Moergeli H, Schnyder U: Outcome of psychiatric treatment: What is relevant for our patients? Compr Psychiatry 45: 199-205 (2004)

28. Haug H J, Stieglitz R D: Das AMDP-System in der klinischen Anwendung und Forschung. Hogrefe, Göttingen, (1997)

29. Haug S, Puschner B, Lambert M J, Kordy H: Veränderungsmessung in der Psychotherapie mit dem Ergebnisfragebogen(EB-45). Zeitschrift für Differentielle und Diagnostische Psychologie 25: 141-151 (2004)

30. Heinrichs D, Hanlon E, Carpenter W: The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. Schizophr Bull 10: 388-398 (1984)

31. Jacobson N S, Follette W C, Revenstorf D: Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. Behaviour Therapy 15: 336-352 (1984)

32. Jacobson N S, Roberts L J, Berns S B, McGlinchey J: Methods for defining and determing the clinical significance of treatment effects: Description, application, and alternatives. J Consult Clin Psychol 67: 300-307 (1999)

33. Jacobson N S, Truax P: Clinical Significance: A statistical approach to defining meaningful change on psychotherapy research. J Consult Clin Psychol 59: 12-19 (1991)

34. Jones S H, Thornicroft G, Coffey M, Dunn G: A brief mental health outcome scale: reliability and validity of the Global Assessment of Functioning (GAF). Br J Psychiatry 166: 654-659 (1995)

35. Kasahara Y: Depression seen in general medical clinics. Psychosom Med 24: 6-14 (1984)

36. Kay S, Opler L, Lindenmayer J: Reliability of positive and negative syndrom scale for schizophrenia. Psychiatry Res 23: 99-110 (1988)

37. Kazarian S S, Joseph L W: A brief scale to help identify outpatients' level of need for community support services. Hosp Community Psych 45: 953-937 (1994)

38. Kessler R C, Barker P R, Colpe L J, Epstein J F: Screening for serious mental Illness in the general opulation. Arch Gen Psychiatry 60: 184-189 (2003)

39. Krieble T: Towards an outcome-based mental health policy for New Zealand. Australas Psychiatry 11: 78-82 (2003)

40. Lambert M J, Burlingame G M, Umphress V, Hansen N B, Vermeersch D A, Clouse G C, Yanchar S C: The reliability and validity of the Outcome Questionnaire. Clin Psychol Psychother 3: 106-116 (1996)

41. Lambert M J, Hannöver W, Nisslmüller K, Richard M, Kordy H: Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). Z Klin Psychol Psychother 31: 40-47 (2002)

42. Lambert M J, Hatch D R, Kingston M D, Edwards B C: Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. J Consult Clin Psych 54: 54-59 (1986)

43. Lambert M J, Okiishi J C, Finch A E, Johnson L D: Outcome assessment: From conceptualization to implementation. Prof Psychol Res Pract 29: 63-70 (1998)

44. Löwe B, Spitzer R L, Zipfel S, Herzog W: Gesundheitsfragebogen für Patienten (PHQ-D), Zweite Auflage. Pfizer, Karlsruhe, (2002)

45. Malewski P: Eine Neufassung des reliable-change Index mit einer Anwendung in der Essstörungsforschung. Dissertation Dr. rer. nat., Technische Universität Carolo-Wilhelmina, Braunschweig (2004)

46. Marks I: Overcoming obstacles to routine outcome measurement. Br J Psychiatry 173: 281-286 (1998)

47. Mestel R, Neeb K, Hauke B, Klingelhöfer J, Stauss K: Zusammenhänge zwischen der Therapiezeitverkürzung und dem Therapieerfolg bei depressiven Patienten. In: Bassler M (Hrsg) Wirkfaktoren von stationärer Psychotherapie, Psychosozial Verlag, Gießen, S. 98-146 (2000)

48. Miller R C, Berman, J S: The efficacy of cognitive behaviour therapies: A quantitative review of the research evidence. Psychol Bull 94, 39-53 (1983)

49. Moessner M, Gallas C, Haug S, Kordy H: The Clinical Psychological Diagnostic System (KPD-38): Sensitivity to change and validity of a self-report instrument for outcome monitoring and quality assurance. Clin Psychol Psychother 18: 331-338 (2011)

50. Moleiro C, Beutler L E: Clinically significant change in psychotherapy for pressive disorders. J Affect Disord 115: 220-224 (2009)

51. Moos R H, Nichol A C, Moos B S: Global Assessment of Functioning ratings an the allocation and outcomes of mental health services. Psychiatr Serv 53: 730-737 (2002)

52. Nilsson E, Wenemark M, Bendtsen P, Kristenson M: Respondent satisfaction regarding SF-36 and EQ-5D, and patients' perspectives concerning health outcome assessment within routine health care. Qual Life Res 16: 1647-1654 (2007)

53. Oliver J, Huxley P, Priebe S: Measuring the quality of life severely mentally ill people using the Lancashire Quality of Life Profile. Soc Psychiatry Psychiatr Epidemiol 32: 76-83 (1997)

54. Oliver J P: The social care directive: development of a quality of life profile for use in community services for the mentally ill. Soc Work Soc Sci Rev 3: 4-45 (1991)

55. Orrell M, Yard P, Handysides J, Schapira R: Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. Br J Psychiatry 174: 409-412 (1999)

56. Overall J E, Gorham D R: The Brief Psychiatric Rating Scale. Psychol Rep 10: 799-812 (1962)

57. Parabiaghi A, Barbato A, D'Avanzo B, Erlicher A, Lora A: Assessing reliable and clinically significant change on Health of the Nation Outcome Scales: Method for displaying longitudinal data. Aust N Z J Psychiatry 39: 719-725 (2005)

58. Phelan M, Slade M, Thornicroft G, Dunn G, Holloway F, Wykes T, Strathdee G, Loftus L, McCrone P, Hayward P: The Camberwell Assessment of Need: The validity and reliability of an instrument to assess the needs of people with severe mental illness. Br J Psychiatry 167: 589-595 (1995)

59. Prel J B du, Hommel G, Röhrig B, Blettner M: Konfidenzintervall oder p-Wert? Teil 4 der Serie zur Bewertung wissenschaftlicher Publikationen. Dtsch Arztebl Int 19: 335-339 (2009)

60. Priebe S, Huxley P, Knight S, Evans S: Application of the Manchester Short Assessment of Quality of Life (MANSA). Int J Soc Psychiatry 45: 7-12 (1999)

61. Prusoff B, Klerman G, Paykel E: Concordance between clinical assessments and patients' self-report in depression. Arch Gen Psychiatry 25: 546-552 (1972)

62. Puschner B, Angermeyer M, Leese M, Thornicroft G, Schene A, Kikkert M, Burti L, Tansella M, Becker T: Course of adherence to medication and quality of life in people with schizophrenia. Psychiatry Res 165: 224-233 (2009)

63. Puschner B, Born A, Giebler A, Helm H, Leese M, Bindman J, Gray R, Schene A, Kikkert M, Burti L, Marrella G, Becker T: Adherence to medication and quality of life in people with schizophrenia. Results of a european multicenter study. J Nerv Ment Dis 194: 746-752 (2006)

64. Puschner B, Schofer D, Knaup C, Becker T: Outcome management in in-patient psychiatric care. Acta Psychiatr Scand 120: 308-319 (2009)

65. Roick C, Kilian R, Matschinger H, Mory C, Angermeyer M C: Die deutsche Version des Client Sociodemographic and Service Receipt Inventory: Ein Instrument zur Erfassung psychiatrischer Versorgungskosten. Psychiatr Prax 28: 84-90 (2001)

66. Ruggeri M, Bisoffi G, Lasalvia A, Amaddeo F, Bonetto C, Biggeri A: A longitudinal evaluation of two-year outcome in a community-based mental health service using graphical chain models. The South-Verona Outcome Project 9. Int J Methods Psychiatr Res 13: 10-23 (2004)

67. Ruggeri M, Lasalvia A, Tansella M, Bonetto C, Abate M, Thornicroft G, Allevi L, Ognibene P: Heterogeneity of outcomes in schizophrenia. Br J Psychiatry 184: 48-57 (2004)

68. Rutledge T, Loh C: Effect Sizes and Statistical Testing in the Determination of Clinical Significance in Behavioral Medicine Research. Ann Behav Med, 27: 138-145(2004)

69. Sainfort F, Becker M, Diamond R: Judgments of quality of life of individuals with severe mental disorders: Patient self-report versus provider perspectives. Am J Psychiatry 153: 492-502 (1996)

70. Salvi G, Leese M, Slade M: Routine use of mental health outcome assessments: Choosing the measure. Br J Psychiatry 186: 146-152 (2005)

71. Schneider F, Härter M, Kratz S, Bermejo I, Mulert C, Hegerl U, Gaebel W, Berger M: Unzureichender subjektiver Behandlungsverlauf bei depressiven Patienten in der haus- und nervenärztlichen Praxis. Z Arztl Fortbild Qualitatssich 97: 57-66 (2003)

72. Shrout P E, Fleiss J L: Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86: 420-428 (1979)

73. Simhandl C, Rogan M, Lesch O M, Musalek M, Strobl R: Wertigkeit von Fremd- und Selbstbeurteilungsskalen bei chronisch Schizophrenen. The value of expert and self-rating scales for chronic schizophrenics. Nervenarzt 55: 371-377 (1984)

74. Slade M: Routine outcome assessment in mental health services. Psychol Med 32: 1339-1343 (2002)

75. Slade M: What outcomes to measure in routine mental health services, and how to assess them: A systematic review. Aust N Z J Psychiatry 36: 743-753 (2002)

76. Slade M, Powell R, Rosen A, Strathdee G: Threshold Assessment Grid (TAG): The development of a valid and brief scale to assess the severity of mental illness. Soc Psychiatry Psychiatr Epidemiol 35: 78-85 (2000)

77. Stieglitz R, Freyberger H J: Selbst- und Fremdbeurteilungsverfahren in der allgemeinen und störungsspezifischen Diagnostik in der Psychotherapie. Self-ratings and ratings by others in general and specific diagnosis in psychotherapy. In: Laireiter, Anton-Rupert Diagnostik in der Psychotherapie. Springer, Wien, S. 295-304 (2000)

78. Süllwold L: Manual zum Frankfurter Beschwerde-Fragebogen (FBF). 1. Auflage, Springer, Berlin, (1991)

79. Szecsenyi J, Stock J, Chenot R, Broge B: QISA – Das Qualitätsindikatorensystem für die ambulante Versorgung. Gesundheits- und Sozialpolitik 2010: 55-61

80. Torrance G: Utility approach to measuring health-related quality of life. J Chronic Dis 40: 593-600 (1987)

81. Trauer T: Consumer and service determinants of completion of a consumer self-rating outcome measure. Australas Psychiatry 12: 48-54 (2004)

82. Trauer T, Callaly T: Concordance between patients and their case managers using the Health of the Nation Outcome Scales (HoNOS). Australas Psychiatry 10: 24-28 (2002)

83. Ueki H, Holzapfel C, Washino K, Inoue M, Ogawa N, Furukawa T: Concordance between self- and observer-ratings on Kasahara's Inventory for the Melancholic Type Personality. Psychiatry Clin Neurosci 56: 569-574 (2002)

84. Voruganti L N, Awad A G, Oyewumi L K, Cortese L, Zirul S, Dhawan R: Assessing health utilities in schizophrenia. A feasibility study. Pharmacoeconomics 17: 273-286 (2000)

85. Ware J E, Sherbourne C D: The MOS 36-item short form health survey (SF-36): Conceptual framework and item selection. Med Care 30: 473-483 (1990)

86. Weinmann S, Puschner B, Becker T: Innovative structuring in German care for schizophrenics. Nervenarzt, 80: 31, 33-34, 36-39 (2009)

87. Weinstein M C, Berwick D M, Goldman P A, Murphy J M, Barsky A J: A comparison of three psychiatric screening tests using receiver operating characeristics (ROC) analysis. Med Care 27: 593-607 (1989)

88. Weissmann M M: The assessment of social adjustment: A review of techniques. Arch Gen Psychiatry 32: 357-365 (1975)

89. WHO: Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F), 2. Auflage, Huber, Bern Göttingen Toronto Seattle, (1993)

90. Wing J K, Beevor A S, Curtis R H, Park S B G, Hadden S: Health of the Nation Outcome Scale (HoNOS). Br J Psychiatry 172: 11-18 (1998)

91. Wise E A: Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. J Pers Assess 82: 50-59 (2004)

92. Wykes T, Stuart E: The measurement of social behaviour in psychiatric patients: an assessment of the reliability and validity of the SBS schedule. Br J Psychiatry 148: 1-11(1996)

93. Zorn U O G: Assessment of quality in healthcare: a universal development scheme for quality indicators. Z Ärztl Fortbild Qualsich 93: 123-128 (1999)

94. Zung W W K: A self-rating depression scale. Arch Gen Psychiatry 12: 63-70 (1965)

7 Anhang

Tabellenverzeichnis

Tabelle 1: Untersuchungen mit mehreren Beurteilerperspektiven 15
Tabelle 2: Cut-Offs und D_{krit}-Werte 32
Tabelle 3: Beschreibung der Stichprobe 35
Tabelle 4: Kennwerte der Instrumente zu T_1 36
Tabelle 5: Einteilung in dysfunktional und funktional zu T_1 36
Tabelle 6: Korrelationen (Pearson) der Instrumente zu T_1 39
Tabelle 7: Kennwerte der Instrumente für T_2 40
Tabelle 8: Einteilung in dysfunktional und funktional zu T_2 41
Tabelle 9: Korrelationen (Pearson) der Instrumente zu T2 42
Tabelle 10: Gegenüberstellung der Mittelwerte (MW) und T-Test 43
Tabelle 11: T-Tests der Unterskalen von HoNOS und EB-45 43
Tabelle 12: Angepasste Partialkorrelationen bei Aufnahme (T_1) 44
Tabelle 13: Angepasste Partialkorrelationen bei Entlassung (T_2) 46
Tabelle 14: Angepasste Partialkorrelationen über die Zeit 46
Tabelle 15: Chi-Quadrat-Test zur Einteilung in funktional und dysfunktional zu T_1 und T_2 48
Tabelle 16: Reliable Veränderung und Gesundung 49
Tabelle 17: Übereinstimmende Bewertungen an Gesundeten 51

Abbildungsverzeichnis

Abbildung 1: Diagramm zur klinischen Signifikanz modifiziert nach [45] 32
Abbildung 2: Histogramme zu den Instrumenten mit Normalverteilungskurven 37
Abbildung 3: Vergleich der Bewertungen des psychischen Zustands zwischen Behandlern, Unabhängigen und Patienten (Spezifität/ Sensitivität) zu T_1 38
Abbildung 4: Vergleich der Einschätzungen von Behandlern, Unabhängigen und Patienten (Spezifität/ Sensitivität) zu T_2 41
Abbildung 5: Streudiagramme zu den Korrelationen der Summenwerte 45
Abbildung 6: „Graphical Chain Modell“ 47
Abbildung 7: Relative Häufigkeiten der als dysfunktional und funktional eingestuften Teilnehmer für HoNOS total score, EB-45 TOT und GAF 48
Abbildung 8: Scatterplot zur klinischen Signifikanz für EB-45 TOT 50
Abbildung 9: Scatterplot zur klinischen Signifikanz für HoNOS total score 50
Abbildung 10: Scatterplot zur klinischen Signifikanz für GAF 51

Fragebögen

1. GAF: Global Assessment of Functioning

Die psychischen, sozialen und beruflichen Funktionen sind auf einem hypothetischen Kontinuum von psychischer Gesundheit bis Krankheit gedacht. Funktionsbeeinträchtigungen aufgrund von körperlichen (oder umgebungsbedingten) Einschränkungen sollten nicht einbezogen werden.

Code (Beachte: Benutze auch entsprechende Zwischenwerte, z.B. 45, 68 oder 72)

100–91 Hervorragende Leistungsfähigkeit in einem breiten Spektrum von Aktivitäten; Schwierigkeiten im Leben scheinen nie außer Kontrolle zu geraten; wird von anderen wegen einer Vielzahl positiver Qualitäten geschätzt; keine Symptome.

90–81 Keine oder nur minimale Symptome (z.B. leichte Angst vor einer Prüfung), gute Leistungsfähigkeit in allen Gebieten, interessiert und eingebunden in ein breites Spektrum von Aktivitäten, sozial effektiv im Verhalten, im allgemeinen zufrieden mit dem Leben, übliche Alltagsprobleme oder –sorgen (z.B. nur gelegentlicher Streit mit einem Familienmitglied).

80–71 Wenn Symptome vorliegen, sind diese vorübergehende oder zu erwartende Reaktionen auf psychosoziale Belastungsfaktoren (z.B. Konzentrationsschwierigkeiten nach einem Familienstreit); höchstens leichte Beeinträchtigung der sozialen, beruflichen und schulischen Leistungsfähigkeit (z.B. zeitweises Zurückbleiben in der Schule).

70–61 Einige leichte Symptome (z.B. depressive Stimmung oder leichte Schlaflosigkeit) ODER einige leichte Schwierigkeiten hinsichtlich der sozialen, beruflichen oder schulischen Leistungsfähigkeit (z.B. gelegentliches Schuleschwänzen oder Diebstahl im Haushalt), aber im allgemeinen relativ gute Leistungsfähigkeit, hat einige wichtige zwischenmenschliche Beziehungen.

60–51 Mäßig ausgeprägte Symptome (z.B. Affektverflachung, weitschweifige Sprache, gelegentliche Panikattacken) ODER mäßig ausgeprägte Schwierigkeiten bezüglich der sozialen, beruflichen oder schulischen Leistungsfähigkeit (z.B. wenige Freunde, Konflikte mit Arbeitskollegen, Schulkameraden oder Bezugspersonen).

50–41 Ernste Symptome (z.B. Suizidgedanken, schwere Zwangsrituale, häufige Ladendiebstähle) ODER jedwede ernste Beeinträchtigung der sozialen, beruflichen und schulischen Leistungsfähigkeit (z.B. keine Freunde; Unfähigkeit, eine Arbeitsstelle zu behalten).

40–31 Einige Beeinträchtigungen in der Realitätskontrolle oder der Kommunikation (z.B. Sprache zeitweise unlogisch, unverständlich oder belanglos) ODER starke Beeinträchtigung in mehreren Bereichen, z.B. Arbeit oder Schule, familiäre Beziehungen, Urteilsvermögen, Denken oder Stimmung (z.B. ein Mann mit einer Depression vermeidet Freunde, vernachlässigt seine Familie und ist unfähig zu arbeiten; ein Kind schlägt häufig jüngere Kinder, ist zu Hause trotzig und versagt in der Schule).

30–21 Das Verhalten ist ernsthaft durch Wahnphänomene oder Halluzinationen beeinflusst ODER ernsthafte Beeinträchtigung der Kommunikation und des Urteilsvermögens (z.B. manchmal inkohärent, handelt grob inadäquat, starkes Eingenommensein von Selbstmordgedanken) ODER Beeinträchtigung der Leistungsfähigkeit in fast allen Bereichen (z.B. bleibt den ganzen Tag im Bett, hat keine Arbeit, kein Zuhause und keine Freunde).

20–11 Selbst- oder Fremdgefährdung (z.B. Selbstmordversuche ohne eindeutige Todesabsicht, häufig gewalttätig, manische Erregung) ODER ist gelegentlich nicht in der Lage, die geringste persönliche Hygiene aufrechtzuerhalten (z.B. schmiert mit Kot) ODER grobe Beeinträchtigung in der Kommunikation (größtenteils inkohärent oder stumm).

10–1 Ständige Gefahr, sich oder andere schwer zu verletzen (z.B. wiederholte Gewaltanwendung) ODER anhaltende Unfähigkeit, die minimale persönliche Hygiene aufrechtzuerhalten ODER ernsthafter Selbstmordversuch mit eindeutiger Todesabsicht.

0 Unzureichende Informationen.

Der aktuelle Zustand des Patienten entspricht einem GAF-Wert von ☐ **(gaf)**

2. EB-45: Ergebnisfragebogen 45

Datum		Code Patient		Fragebogen-Nr.	

Sehr geehrte Dame, sehr geehrter Herr,

Uns würde interessieren, wie Sie sich in der letzten Woche (einschließlich heute) gefühlt haben. Bitte lesen Sie sich sorgfältig jede Frage durch und kreuzen das Kästchen derjenigen Kategorie an, welche am ehesten auf Ihre gegenwärtige Situation zutrifft. Unter den Begriff 'Arbeit' fallen alle Tätigkeiten wie reguläre Arbeitsstelle, Schule, Ausbildung, Hausarbeit, ehrenamtliche Tätigkeiten usw.

		nie	selten	manchmal	häufig	fast immer
1.	Ich komme gut mit anderen aus.	☐	☐	☐	☐	☐
2.	Ich werde schnell müde.	☐	☐	☐	☐	☐
3.	Ich fühle mich interessenlos.	☐	☐	☐	☐	☐
4.	Ich fühle mich von der Arbeit/Ausbildung gestresst.	☐	☐	☐	☐	☐
5.	Ich gebe mir selbst die Schuld für bestimmte Dinge.	☐	☐	☐	☐	☐
6.	Ich fühle mich gereizt.	☐	☐	☐	☐	☐
7.	Ich fühle mich unglücklich in meiner Ehe/Partnerschaft.	☐	☐	☐	☐	☐
8.	Ich habe Gedanken, mir das Leben zu nehmen.	☐	☐	☐	☐	☐
		nie	selten	manchmal	häufig	fast immer
9.	Ich fühle mich schwach.	☐	☐	☐	☐	☐
10.	Ich bin besorgt und ängstlich.	☐	☐	☐	☐	☐
11.	Wenn ich zuviel getrunken habe, muss ich am nächsten Morgen wieder trinken, um auf die Beine zu kommen. (*Wenn Sie nicht trinken, kreuzen Sie bitte "nie" an.*).	☐	☐	☐	☐	☐
12.	Ich bin zufrieden mit meiner beruflichen Situation/der Ausbildung.	☐	☐	☐	☐	☐
13.	Ich bin ein glücklicher Mensch.	☐	☐	☐	☐	☐
14.	Ich arbeite/lerne zu viel.	☐	☐	☐	☐	☐
15.	Ich fühle mich wertlos.	☐	☐	☐	☐	☐
16.	Ich bin wegen Ärger in der Familie beunruhigt.	☐	☐	☐	☐	☐
17.	Ich habe eine unerfüllte Sexualität.	☐	☐	☐	☐	☐
18.	Ich fühle mich einsam.	☐	☐	☐	☐	☐
19.	Ich streite mich häufig.	☐	☐	☐	☐	☐
		nie	selten	manchmal	häufig	fast immer
20.	Ich fühle mich geliebt und erwünscht.	☐	☐	☐	☐	☐

21.	Ich genieße meine Freizeit.	☐	☐	☐	☐	☐
22.	Ich habe Schwierigkeiten, mich zu konzentrieren.	☐	☐	☐	☐	☐
23.	In Bezug auf die Zukunft fühle ich mich hoffnungslos.	☐	☐	☐	☐	☐
24.	Ich mag mich.	☐	☐	☐	☐	☐
25.	Mir kommen störende Gedanken, die ich nicht loswerden kann.	☐	☐	☐	☐	☐
26.	Ich ärgere mich über Leute, die mich wegen des Trinkens / Gebrauchs von Drogen kritisieren. (*Wenn diese Frage auf Sie nicht zutrifft, kreuzen Sie bitte "nie" an.*).	☐	☐	☐	☐	☐
27.	Ich habe einen empfindlichen Magen.	☐	☐	☐	☐	☐
28.	Ich arbeite/lerne nicht mehr so gut wie früher.	☐	☐	☐	☐	☐
29.	Ich habe Herzklopfen.	☐	☐	☐	☐	☐
30.	Ich habe Mühe, mit meinen Freunden oder engeren Bekannten auszukommen.	☐	☐	☐	☐	☐
		nie	selten	manchmal	häufig	fast immer
31.	Ich bin mit meinem Leben zufrieden.	☐	☐	☐	☐	☐
32.	Ich habe Schwierigkeiten bei der Arbeit/in der Ausbildung wegen des Trinkens oder wegen des Gebrauchs von Drogen. (*Wenn diese Frage auf Sie nicht zutrifft, kreuzen Sie bitte "nie" an.*).	☐	☐	☐	☐	☐
33.	Ich habe das Gefühl, dass etwas Schlimmes passieren wird.	☐	☐	☐	☐	☐
34.	Ich habe Gliederschmerzen.	☐	☐	☐	☐	☐
35.	Ich habe Angst vor offenen Plätzen, mit öffentlichen Verkehrsmitteln zu fahren, dem Autofahren usw..	☐	☐	☐	☐	☐
36.	Ich bin nervös.	☐	☐	☐	☐	☐
37.	Ich habe erfüllte partnerschaftliche Beziehungen.	☐	☐	☐	☐	☐
38.	Ich habe das Gefühl, dass ich meine Aufgaben bei der Arbeit / in der Ausbildung nicht gut mache.	☐	☐	☐	☐	☐
39.	Ich habe zu viele Meinungsverschiedenheiten bei der Arbeit / in der Ausbildung.	☐	☐	☐	☐	☐
40.	Ich habe das Gefühl, dass etwas mit meinem Verstand nicht in Ordnung ist.	☐	☐	☐	☐	☐
41.	Ich habe Einschlafstörungen oder Durchschlafstörungen.	☐	☐	☐	☐	☐
		nie	selten	manchmal	häufig	fast immer
42.	Ich bin niedergeschlagen.	☐	☐	☐	☐	☐
43.	Mit meinen Beziehungen zu anderen bin ich zufrieden.	☐	☐	☐	☐	☐
44.	Bei der Arbeit / in der Ausbildung bin ich so zornig, dass ich Dinge tun könnte, die mir später leid tun.	☐	☐	☐	☐	☐
45.	Ich habe Kopfschmerzen.	☐	☐	☐	☐	☐

1. HoNOS: Health of the Nation Outcome Scale

HoNOS-D Rating

Beurteilen Sie für jedes der 12 Items das jeweils **schwerste** Problem, das während der **letzten 7 Tage** aufgetreten ist.

Legende: Tragen Sie für jedes Item 0 - 4 oder 9 in die Kästchen ein.

0	kein Problem	**1**	klinisch unbedeutendes Problem ohne Handlungsbedarf
2	leichtes Problem, aber eindeutig vorhanden	**3**	eher schweres Problem
4	schweres bis sehr schweres Problem	**9**	nicht bekannt / nicht anwendbar

1) Überaktives, aggressives, Unruhe stiftendes oder agitiertes Verhalten ☐

2) Absichtliche Selbstverletzung ☐

3) Problematischer Alkoholkonsum oder Drogenkonsum ☐

4) Kognitive Probleme ☐

5) Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung ☐

6) Probleme in Zusammenhang mit Halluzinationen und Wahnvorstellungen ☐

7) Gedrückte Stimmung ☐

8) Andere psychische und verhaltensbezogene Probleme ☐

Falls andere psychische /verhaltensbezogene Probleme vorhanden, geben Sie die Art der Probleme an

A Phobisch — *B Angst*
C Zwangsgedanken/ -handlungen — *D Psychische Belastung / Anspannung* ☐
E Dissoziativ — *F Somatoform*
G Essen — *H Schlaf*
I Sexuell — *J Andere (Spezifizieren)* _____

9) Probleme mit Beziehungen ☐

10) Probleme mit alltäglichen Aktivitäten ☐

11) Probleme durch die Wohnbedingungen ☐

12) Probleme durch die Bedingungen im Beruf und im Alltag ☐

8 Danksagungen

Mein Dank gilt Herrn Prof. Dr. Thomas Becker für die Überlassung des Themas.
Besonders danke ich auch Herrn PD Dr. Bernd Puschner für die freundliche und geduldige Betreuung über die Jahre und für die wertvollen Hinweise bei der Durchsicht dieser Arbeit.
Bei Herrn Prof. Dr. Jörn von Wietersheim bedanke ich mich herzlich für die Übernahme des Zweitgutachtens.
Ebenfalls danke ich Frau Dr. Piroska Luka für die schnelle und sorgfältige Rechtschreibkorrektur.
Großer Dank gebührt auch meinen Eltern und Geschwistern, die mich immer unterstützen und einen wichtigen Platz in meinem Leben einnehmen. Vor allem mein Bruder hat mir mit seinem fachkundigen Rat sehr geholfen.
Nicht zuletzt bedanke ich mich auch bei meinem Ehemann für sein Verständnis und seine Ratschläge und bei meiner Tochter Elisabeth, die mich in den letzten Monaten der Fertigstellung dieser Arbeit mit ihrem Lächeln stets wieder motivierte.

9 Lebenslauf

Lebenslauf aus Gründen des Datenschutzes entfernt.