

Zusammenfassung von Studien

„Zur Fremdeinschätzung des Schweregrades psychischer Erkrankungen: Die Deutsche Version der Health of the Nation Outcome Scales, HoNOS-D“

Assoz. Prof. Dr. phil. Dipl.-Psych. Sylke Andreas, M.A.

Stand Mai 2010

Inhaltsverzeichnis

1. Einleitung	3
2. Zentrale Instrumente zur Fremdeinschätzung der Symptomschwere	7
3. Die deutsche Version der Health of the Nation Outcome Scales - HoNOS	12
3.1. Der Übersetzungsprozess der HoNOS in die deutsche Sprache.....	12
3.2. Aufbau der HoNOS-D	12
4. Psychometrische Eigenschaften der HoNOS: Theoretischer Hintergrund	15
4.1. Zur Praktikabilität der HoNOS	15
4.2. Zur Reliabilität der HoNOS	15
4.3. Zur Validität der HoNOS.....	17
5. Eigene Studien	21
5.1. Kontext und Begründung.....	21
5.2. Ziele und Fragestellungen.....	21
5.3. Design und Methodik.....	22
5.4. Stichprobenbeschreibung.....	24
5.5. Verwendete Instrumente	25
5.6. Statistische Analysen	29
6. Ergebnisse	36
6.1. Praktikabilität.....	36
6.2. Itemkennwerte der HoNOS-D	37
6.3. Reliabilität.....	40
6.3.1 Test-Retest-Reliabilität	40
6.3.2 Interraterreliabilitäten	40
6.4. Validität	44
6.4.1 Faktorielle Validität.....	44
6.4.2 Konvergente und diskriminante Validität.....	45
6.4.3 Kriteriumsvalidität.....	47
6.4.5 Änderungssensitivität	48
6.4.6 Prognostische Validität.....	51
7. Zusammenfassung	53
8. Implikationen für Versorgung und Forschung	59
8.1. Klinische Implikationen.....	59
8.2. Forschungsbezogene Implikationen	60
9. Literatur	62

1. Einleitung

Kontext und Zielsetzung der Studien

Die **differenzierte Erfassung des Schweregrades** ist ein zentraler Bestandteil von vielen verschiedenen Fragestellungen in der Diagnostik und Behandlung von Patienten mit psychischen Störungen. So nimmt sie zum Beispiel eine entscheidende Rolle in der Zuweisung von Patienten zu bestimmten Therapieangeboten, z.B. von Patienten mit einer schweren Borderline-Persönlichkeitsstörung zur Dialektisch-Behavioralen Therapie von Linehan (1996) oder zu spezialisierten Einrichtungen, ein. Die korrekte Diagnose und damit verbunden die richtige Einordnung der Schwere einer Erkrankung spielt somit auch bei der Therapieplanung eine entscheidende Rolle. Auch bei der Ressourcenvergabe bzw. -planung bei Patienten in stationärer psychotherapeutischer Behandlung ist eine Erfassung des Schweregrades von Bedeutung und kann auch mit Ressourcenallokation durch Kostenträger verbunden sein. Nicht zuletzt ist die Erfassung der Schwere einer psychischen Erkrankung auch direkt mit der Messung des Therapieerfolges und damit mit Qualitätssicherungsmaßnahmen assoziiert. Zielsetzung sollte es deswegen sein, Instrumentarien zur Verfügung zu stellen, anhand derer eine zuverlässige und valide Einschätzung des Schweregrades vorgenommen werden kann.

In den letzten Jahrzehnten wurden deswegen vielfältige Bemühungen unternommen, reliable und valide Instrumente zur Erfassung des Schweregrades einer psychischen Störung zu entwickeln. Die Schwere einer psychischen Erkrankung kann mittels **Selbst- oder Fremdeinschätzung** vorgenommen werden. Selbstbeurteilungsverfahren werden in der klinischen Praxis und Forschung aufgrund verschiedener Vorteile häufiger eingesetzt als Fremdbeurteilungsverfahren. So ist ein Großteil klinisch relevanter Informationen nur über Selbstauskunft zu erheben (z.B. die Stimmung oder Befindlichkeit eines Patienten), zudem sind sie zeitökonomisch und in sehr kurzen Zeitabständen einsetzbar. Die meisten gängigen Verfahren sind teststatistisch überprüft, verfügen partiell aber über Schwächen in ihren psychometrischen Gütekriterien (z.B. keine ausreichenden Angaben zur Reliabilität, meist unbefriedigende Angaben zur Validität oder Fehlen von Normwerten). Dagegen sind Fremdeinschätzungsinstrumente oft von Vorteil, wenn aufgrund der Komplexität der Situation der Patient selbst keine Auskunft geben könnte. Sie ermöglichen einen anderen Blick auf die Situation und können damit zum zusätzlichen Informationsgewinn beitragen. Trotz ihrer Vorteile weisen auch Fremdbeurteilungsver-

fahren einige Nachteile auf: So sind sie z.B. zeitaufwendiger, da sie oft mit Trainings verbunden sind und können mit Erinnerungs- und Gedächtnisfehlern während der Einschätzung einhergehen. Daraus kann geschlussfolgert werden, dass Selbst- und Fremdeinschätzung eine komplementäre, sich ergänzende Funktion zukommt. Obwohl eine differenzierte Erfassung des Schweregrades, insbesondere aus der Perspektive der Fremdeinschätzung, aus den dargelegten Gründen erforderlich ist, gibt es in Deutschland kaum differenzierte und psychometrisch gut überprüfte Instrumente.

Das **Hauptziel** der vorliegenden Arbeit war es demzufolge, ein international weit verbreitetes Fremdeinschätzungsinstrument zur differenzierten Erfassung der Schwere einer psychischen Störung („Health of the Nation Outcome Scales, HoNOS“) in die deutsche Sprache zu übersetzen und psychometrisch zu überprüfen. Die psychometrische Überprüfung wurde an verschiedenen großen und weitestgehend klinisch repräsentativen Stichproben aus dem Bereich der stationären psychosomatischen/psychotherapeutischen Versorgung sowie der stationären Behandlung Suchtkranker durchgeführt. Es wurde außerdem eine Weiterentwicklung eines Items des Fremdeinschätzungsinstrumentes von unserer Arbeitsgruppe vorgenommen, welche ebenfalls anhand einer großen klinischen Stichprobe von Patienten eingängiger psychometrischer Testung unterzogen wurde. Schließlich interessierte uns auch die Überprüfung der Eignung des Schweregrades innerhalb der Stellung eines Klassifikationssystems zum Ressourcenverbrauch bei Patienten in stationärer psychotherapeutischer Behandlung.

Die Studien wurden im Rahmen eines von der DFG-geförderten Forschungsprojektes „Überprüfung der Validität eines Fremdratings zur differenzierten Erfassung des Schweregrades: „Health of the Nation Outcome Scales“ (GZ AN 382/3-1) und im Rahmen eines vom Norddeutschen Verbund für Rehabilitationsforschung (NVRF) initiierten Projektes „Entwicklung eines Klassifikationssystems für Patienten mit psychischen Erkrankungen“, welches vom BMBF und der Deutschen Rentenversicherung Bund finanziert wurde (01GD 9808/0), durchgeführt.

Danksagung

Die zahlreichen Studien waren nur möglich, weil mich viele Kollegen und Kolleginnen unterstützt haben. Zunächst möchte ich mich für ihren steten großartigen Einsatz und ihr Engagement bei der Datenerhebung und –organisation im HoNOS-D Projekt bei Maria Hausberg bedanken. David Horstmann danke ich recht herzlich für seine Unterstützung in der Pflege und Aufbereitung der verschiedenen Datensätze. Mein besonderer Dank gilt auch den Diplomandinnen Marlies Johst und Pia Theisen, die durch die Einschätzungen im HoNOS-D dazu beigetragen haben, noch mehr notwendige und hilfreiche Erfahrungen mit dem Instrument zu sammeln. Sven Rabung, Stephan Kawski, Timo Harfst, Jörg Dirmaier und Christiane Bleich möchte ich danken für ihren sehr großen Einsatz in der Gewinnung der Daten und in der unglaublichen Geduld bei der Koordination des Gesamtprojektes des Qualitätssicherungsprogramms der Gesetzlichen Krankenkassen in der medizinischen Rehabilitation (QS-Reha[®]). Mein spezifischer Dank gilt besonders Timo Harfst und Jörg Dirmaier, die an der Übersetzung des Instrumentariums mitgewirkt haben. Anna Levke Brütt und Jana Volkert möchte ich danken für ihre Diskussionen um inhaltliche Fragestellungen und ihre besondere tatkräftige Unterstützung in der Koordination und Bewältigung der zahlreichen Aufgaben in unserer Arbeitsgruppe. Ebenso gilt mein Dank Kollegen und Kolleginnen, die in anderen, gemeinsamen Forschungsprojekten mitgearbeitet haben und somit auch zum Gelingen dieser Arbeit beigetragen haben: Matthias Schmitz, Ulrich Prinz und Olga Wlodarczyk.

Ganz herzlich möchte ich mich bei Holger Schulz bedanken, der mir ein sehr steter Wegbegleiter ist und mich tatkräftig bei der Verfolgung von Projektideen und Projektzielen unterstützt. Er hat mich darüber hinaus auf meinen Weg zur Habilitation kontinuierlich gefördert und es verstanden durch anregende Diskussionen und Überlegungen mein Forschungsinteresse noch mehr zu wecken. Ebenso möchte ich sehr gern Uwe Koch für seine Impulssetzungen und entscheidenden Hinweise zur Entstehung der Projektidee danken. Durch ihn war es möglich, einen Kontakt zum Royal College of Psychiatrists in London herzustellen und die Erlaubnis für die Übersetzung des Instrumentes einzuholen. Ich möchte mich auch bei Martin Härter für sein Engagement insbesondere in der Schlussphase der Habilitation bedanken und dafür, dass er mich in einem weiteren gemeinsamen Forschungsprojekt sehr unterstützt. Ebenfalls möchte ich mich für die Begleitung und Förderung auf meinem Weg zur Habilitation bei Monika Bullinger bedanken. Ich möchte auch die Gelegenheit nutzen und mich an dieser Stelle für eine ganz besondere, inzwischen zur Freundschaft gewachsene Beziehung bedanken:

mein Dank gilt Isa Sammet, welche mich in meinen allerersten Forschungsbemühungen und –interessen motivierend unterstützt und mich auf meinem Weg bis zur Habilitation stets sehr engagiert begleitet hat. Mein ganz besonderer Dank gilt an dieser Stelle Sven Rabung, der mir sowohl fachlich als auch emotional ein wichtiger Wegbegleiter ist. Bei meinen Eltern Marlies und Günter Andreas möchte ich mich bedanken, dass sie mich immer bei meiner beruflichen Zielerreichung sehr unterstützt haben und stets geduldig an meiner Seite waren.

Die zahlreichen Studien und Datenerhebungen waren möglich, weil mich viele externe Kooperanden unterstützt haben. An dieser Stelle sind alle am Qualitätssicherungsprogramm der Gesetzlichen Krankenkassen (QS-Reha[®]) beteiligten Einrichtungen dankend zu nennen. Ebenfalls möchte ich mich sehr herzlich bei Karin Harries-Hedder, Dieter Adamski, Wolfgang Schwenk und Henning Schauenburg für die Kooperationsmöglichkeit und den engagierten Einsatz des Instrumentes in ihren Einrichtungen bedanken. Mein besonderer Dank gilt auch Robert Mestel, der mir durch das Zurverfügungstellen von diagnostischen Videos die Möglichkeit gegeben hat, hilfreiche und notwendige Erfahrungen mit dem Instrument zu sammeln. In diesem Zusammenhang gilt mein Dank auch allen Patienten, die sich bereit erklärt haben an den Studien teilzunehmen.

Gliederung der Arbeit

Die vorliegende Arbeit gliedert sich in 8 Kapitel. Sie beginnt mit einer Vorstellung von zentralen und weit verbreiteten Instrumenten zur Fremdeinschätzung des Schweregrades (Kap. 2). In Kap. 3 wird der Übersetzungsprozess und das Instrument vorgestellt. In Kap. 4 folgt eine Darstellung empirischer Befunde zum HoNOS anderer Studien. In Kap. 5 werden Kontext, Fragestellungen, Design, Stichprobenbeschreibung und statistische Auswertungsverfahren der durchgeführten eigenen Studien beschrieben. Die Ergebnisse der für die Habilitation relevanten Einzelarbeiten, die im Anhang zu finden sind, werden in Kap. 6 kurz dargestellt (Anhang 3, 4, 5, 6 und 7). In Kap. 7 werden die Befunde kritisch vor dem Hintergrund empirischer Ergebnisse anderer Studien und bezogen auf die Forschungsfragestellungen eingeordnet. In Kap. 8 werden Implikationen für die Praxis und für die Forschung abgeleitet.

2. Zentrale Instrumente zur Fremdeinschätzung der Symptomschwere

Ein wesentlicher Bestandteil der Qualitätssicherung und Evaluation von Versorgungsangeboten im Bereich von psychischen Erkrankungen ist die praktikable und zuverlässige Erfassung des Therapieergebnisses. Wie bereits erwähnt kann die Symptomschwere mittels Selbsteinschätzung, also durch den Patienten¹ selbst, oder durch Fremdeinschätzung, also z.B. durch einen Therapeuten, erfolgen. Um eine spätere Bewertung der empirischen Ergebnisse zur psychometrischen Überprüfung der deutschen Version der „Health of the Nation Outcome Scales, HoNOS“ (Andreas et al., 2007; Wing et al., 1998) auf eine breitere Basis stellen zu können, sollen im Folgenden zentrale und im deutschen Sprachraum weitverbreitete Fremdeinschätzungsinstrumente beschrieben werden.

Tabelle 1 gibt einen Überblick über eine Auswahl an *gut etablierten Fremdeinschätzungsinstrumenten*, die nachfolgend vorgestellt und abschließend unter dem Aspekt psychometrischer Eigenschaften bewertet werden sollen. Das **AMDP System** (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP), 1996) wurde mit der Zielsetzung der Dokumentation psychopathologischer Befunde von der Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie entwickelt (Collegium Internationale Psychiatriae Salarum (CIPS), 1996c). Es handelt sich um ein umfassendes System, welches gleichermaßen soziodemographische Daten, anamnestische Daten, einen psychischen und somatischen Befund des Patienten erhebt. Das AMDP-System enthält, allein für den psychischen Befundteil, 100 einzuschätzende Merkmale. Es lassen sich aus den Symptomen des psychischen und des somatischen Befundes 9 Syndromskalen (z.B. depressives oder manisches Syndrom) sowie 3 übergeordnete Syndrome (paranoid-halluzinatorische Symptomatik, depressive und psychoorganische Symptomatik) bilden. Um eine Beurteilung im AMDP-System vornehmen zu können, sollte der Anwender mit den klinischen Konzepten der Items vertraut sein und über ein Training verfügen. Die Bearbeitungszeit beträgt 30 bis 60 Minuten. Die Reliabilitätskoeffizienten liegen für die meisten Einzelitems eher in einem moderaten bis zufrieden stellenden (0.40 – 0.68), für die Syndromskalen mit Werten zwischen 0.63 – 0.93 in einem zufrieden stellenden

¹ Im Folgenden soll jeweils die grammatikalisch männliche Form verwendet werden. Darin eingeschlossen ist auch die weibliche Form.

Bereich. Aspekte zur diskriminanten Validität hinsichtlich verschiedener Patientengruppen können als gut bewertet werden (Baumann & Stieglitz, 1989). Validitätsnachweise bezüglich des Zusammenhanges mit anderen Fremdbeurteilungsinstrumenten wie der Brief Psychiatric Rating Scale, BPRS (Collegium Internationale Psychiatrica Scalarum (CIPS), 1996a) sind ebenfalls als gut zu bewerten (Baumann & Stieglitz, 1989).

Tabelle 1: Überblick über zentrale Fremdeinschätzungsinstrumente zur Erfassung des Schweregrades einer psychischen Störung (Andreas, 2005)

<i>Verfahren (Autoren)</i>	<i>Zielsetzung</i>	<i>Aufbau</i>	<i>Praktikabili- tät</i>	<i>Reliabili- tät</i>	<i>Validität</i>
AMDP-System (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP), 1996)	Erfassung psychopathologischer Befunde	4 Dimensionen (Soziodemographie, Anamnese, psychischer und somatischer Befund) 140 Items 9 Subskalen 3 übergeordnete Skalen	Bearbeitungszeit: 30 bis 60 Minuten Training erforderlich	Für 68 Items moderat bis zufriedenstellende Werte	Gut ¹
Beeinträchtigungsschwere-Score (BSS, (Schepank, 1995)	Erfassung der Schwere einer psychischen Erkrankung	3 Dimensionen Gesamtwert	Bearbeitungszeit: ca. 2 Minuten	Gut	Gut ¹
Brief Psychiatric Rating Scale (BPRS (Overall & Gorham, 1976)	Erfassung des Schweregrades psychischer Symptome	18 Symptomkomplexe Gesamtwert 5 Subskalen	Bearbeitungszeit: ca. 25 Minuten	Gut	Gut ¹
CGI (National Institute of Mental Health, 1976)	Nutzen-Risiko-Bewertung bei medikamentöser Behandlung	3 Items	k.A.	Moderat	Gut ²
Skala zur Globalen Erfassung des Funktionsniveaus (GAF, (Endicott et al., 1976)	Erfassung des Funktionsniveaus	Gesamtwert	Bearbeitungszeit: ca. 1 Minute	Moderat bis gut	Gut ¹ bis auf Schwächen in der diskriminierenden Validität

Anmerkungen: ¹Bewertung der Gütekriterien erfolgte nach Stieglitz und Freyberger (Stieglitz & Freyberger, 2001), Validität Gut = ausreichend empirische Studien durchgeführt mit guten Ergebnissen; ²Bewertung aufgrund vorliegender empirischer Befunde aus Validitätsstudien zu den einzelnen Instrumenten

Ein international zur Validierung der HoNOS sehr häufig eingesetztes Instrument ist die **Brief Psychiatric Rating Scale**, BPRS (Overall & Gorham, 1976). Die BPRS entstand aus einer Verkürzung der Multidimensional Scale for Rating Psychiatric Patients MSRPP (Lorr et al.,

1953) und der Inpatient Multidimensional Psychiatric Scale, IMPS (Lorr et al., 1963). Overall und Gorham (1976) bildeten auf der Grundlage von Expertenratings und multivariaten Techniken zunächst 16 später dann 18 Symptomkomplexe, die 7-stufig einzuschätzen sind („nicht vorhanden“ bis „extrem stark“). 1986 wurde die BPRS von Hiller et. al. in die deutsche Sprache übersetzt (Hiller et al., 1986). Die Informationen für das Rating werden aus einem ca. 20 Minuten dauernden Interview gewonnen, welches durch eine gezielte Befragung ergänzt wird. Die Bearbeitungsdauer beschränkt sich auf ca. 3 Minuten. Die Reliabilität der BPRS lässt sich größtenteils als gut bewerten: Interraterreliabilitäten für die Faktoren nehmen Werte zwischen 0.52 und 0.90 ein, für den Gesamtwert liegen die Werte zwischen 0.87 und 0.97. Auch die Konstrukt- und kriterienbezogene Validität der BPRS gilt als gut belegt (Collegium Internationale Psychiatriae Salarum (CIPS), 1996a).

Die **Clinical Global Impressions Scale (CGI)** wurde 1970 vom National Institute of Mental Health im Rahmen der Schizophrenieforschung entwickelt (Collegium Internationale Psychiatriae Salarum (CIPS), 1996b). Hauptsächliches Ziel des Instrumentes ist die Nutzen-Risiko-Bewertung bei medikamentöser Behandlung von Patienten mit psychischen Störungen. Die Skala besteht aus 3 Items, wobei die ersten beiden Items in 7-stufiger Ausprägung beurteilt werden sollen: Schweregrad der Krankheit, Gesamtbeurteilung der Zustandsänderung und Therapeutische Wirksamkeit/Unerwünschte Wirkungen (Therapie-Risiken). Eine Besonderheit und Folge einer überarbeiteten Version stellt die Anweisung dar, dass der Beurteiler jeweils den Bezugsrahmen des „typischen Patienten“ und nicht seinen gesamten Erfahrungsschatz zugrunde legen sollte. Die Interraterreliabilitäten für die drei Items der CGI Skala sind in einem eher moderaten Bereich (0.35 bis 0.66), Retest-Reliabilitäten variieren auch stark (Werte zwischen 0.15 und 0.81). Eine Untersuchung zur Validität der CGI Skala zeigte, dass bei Demenzpatienten die Beurteilung des Schweregrades der Erkrankung wesentlich durch diagnostische Kriterien (DSM-III-R) bestimmt wird (Collegium Internationale Psychiatriae Salarum (CIPS), 1996b).

Zusammenfassung

Zwei sowohl in der Forschung als auch in der klinischen Praxis in Deutschland häufig verwendeten Instrumente zur Fremdeinschätzung des Schweregrades ist der **Beeinträchtigungsschwere-Score (BSS)** von Schepank (1995) und die **Skala zur Globalen Erfassung des Funktionsniveaus (GAF)** (Endicott et al., 1976; Saß et al., 1998).

Die **Global Assessment of Functioning Scale** oder **Skala zur Globalen Erfassung des Funktionsniveaus (GAF)** erfasst das allgemeine Funktionsniveau unter Berücksichtigung psychischer, sozialer und beruflicher Aspekte von Patienten mit psychischen Störungen (Saß et al., 1998) und bildet die Achse V des DSM-IV. Die Beurteilung des Ausmaßes der Schwere einer Funktionsbeeinträchtigung wird anhand einer Skala von 1-100 getroffen, wobei ein Wert von 100-91 eine hervorragende Leistungsfähigkeit und ein Wert von 10-1 eine sehr stark beeinträchtigte Leistungsfähigkeit widerspiegelt. Beeinträchtigungen aufgrund körperlicher Erkrankung werden in die Einschätzung nicht miteinbezogen. Obwohl die empirische Basis als noch nicht ausreichend beurteilt werden muss, zeigen doch vorliegende Befunde gute Gütekriterien für die GAF-Skala. Zur Reliabilität der Skala werden Werte zwischen 0.54 und 0.90 berichtet, wobei sich diese große Spannweite durch den Trainings- bzw. Erfahrungsstand der Beurteiler erklären lässt. Eine neuere Studie von Vatnaland et al. (2007) konnte allerdings aufzeigen, dass für Beurteilerübereinstimmungen in einem klinischen Setting nur moderate Reliabilitätskennwerte erzielt werden konnten. Hinsichtlich der Validität können gute Ergebnisse verzeichnet werden: So korreliert die GAF-Skala z.B. mit der WHO-Disability-Diagnostic-scale mit $r = -0.64$ (American Psychiatric Association, 2002).

Ein fester Bestandteil bei der Fremdeinschätzung des Schweregrades in Qualitätssicherungsprogrammen von Patienten mit psychischen Erkrankungen ist der **Beeinträchtigungsschwerescore, BSS**, (Schepank, 1995). Er erhebt wie schwer ein Patient im psychischen, körperlichen und sozialkommunikativen Bereich durch seine Symptomatik beeinträchtigt ist. Es steht ein Profilwert zur Verfügung, der die Bevorzugung der Beeinträchtigung wider gibt und ein Summenwert, der die Gesamtbeeinträchtigung wider spiegelt (Schepank, 1995). Laut dem Testautor wird eine gründliche und persönliche Diagnostik (ca. 1 bis 1,5 h Interview) eines in der Beurteilung psychischer Krankheiten erfahrenen Experten für eine Einschätzung im BSS vorausgesetzt (Schepank, 1995). Der Beurteilungszeitraum kann sich zwischen den letzten 7 Tagen oder den letzten 12 Monaten bewegen, im Gegensatz zur GAF-Skala, bei welcher die Beurteilung zum aktuellen Zeitpunkt vorgenommen wird (Lange & Heuft, 2002). Interraterreliabilitäten können als zufrieden stellend bewertet werden mit Werten zwischen 0.89 und 0.99, für die Subdimensionen zwischen 0.68 und 0.97. Die bislang vorliegenden Studien zur Validität zeigen ebenfalls gute Ergebnisse (Schepank, 2002).

Hinsichtlich der aufgeführten Fremdeinschätzungsinstrumente lässt sich zunächst festhalten, dass einige Messverfahren über eher moderate psychometrische Eigenschaften verfügen (z.B.

das AMDP-System oder der CGI). Ein weiteres Beurteilungskriterium für ein Instrumentarium, welches in der Routineversorgung zuverlässig den Schweregrad einschätzen sollte, ist die Praktikabilität. Bei der Bewertung der Praktikabilität fällt auf, dass kaum eines der aufgeführten Systeme durch in der Routine realisierbare Bearbeitungszeiten gekennzeichnet ist: So z.B. im Fall der BPRS, bei der durchschnittlich 25 Minuten zur Beantwortung benötigt werden. Von den aufgeführten Fremdeinschätzungsinstrumenten kommen deswegen hauptsächlich nur zwei in Betracht: der BSS und der GAF. Es lässt sich allerdings festhalten, dass unabhängig von den psychometrischen Gütekriterien sowohl der BSS als auch der GAF nach erfolgter eingehender *inhaltlicher* Prüfung nicht für eine differenzierte Abbildung des Schweregrades in Frage kommen, da die Operationalisierung der drei verschiedenen Dimensionen des BSS und auch die Ankerpunkte der GAF-Skala sehr unscharf sind.

Deswegen war es Zielsetzung vorliegender Arbeit ein Fremdeinschätzungsinstrument zur differenzierten Erfassung des Schweregrades, welches in der Routineversorgung einsetzbar und in verschiedenen Behandlungssettings geeignet ist, zu identifizieren, in die deutsche Sprache zu übersetzen und psychometrisch zu überprüfen.

3. Die deutsche Version der Health of the Nation Outcome Scales - HoNOS

3.1. Der Übersetzungsprozess der HoNOS in die deutsche Sprache

In Anlehnung an das von der Weltgesundheitsorganisation empfohlene Vorgehen (Sartorius & Kuyken, 1994) wurde im Februar 2003 in einem ersten Schritt das Score-Sheet und das umfangreiche Glossar der HoNOS in einem expertengestützten Konsensusverfahren von drei Mitgliedern (Sylke Andreas, Timo Harfst und Jörg Dirmaier) der Arbeitsgruppe für Psychotherapie- und Versorgungsforschung (AGPV) am Institut und Poliklinik für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf von der englischen in die deutsche Sprache übersetzt. Dieser Schritt erfolgte bereits autorisiert durch die Testautoren der englischen Originalversion des Royal College of Psychiatrist's in London (Großbritannien). Im nächsten Schritt wurde die deutsche Übersetzung von einem unabhängigen Experten (Holger Schulz) - ebenfalls mit langjähriger klinischer Erfahrung - begutachtet. Seine Korrekturen wurden dann mit den drei Übersetzern diskutiert und eingearbeitet. Anschließend wurde diese erste deutsche Version von einem „Native Speaker“ zurückübersetzt und den Testautoren der Originalfassung (vertreten durch Paul Lelliott und Mick James) zum Abgleich mit der englischen Originalversion der HoNOS zur Korrektur übersandt. Nach geringfügigen Veränderungen und Anmerkungen durch die Testautoren erfolgte die Fertigstellung der endgültigen Fassung des Score-Sheets und des Glossars der deutschen Version der HoNOS-D im April 2003 (Andreas et al., 2007).

3.2. Aufbau der HoNOS-D

Die HoNOS-D ist ein 12 Item umfassendes Instrument zur Erfassung des differenzierten Schweregrades einer psychischen Erkrankung (siehe Anhang 1). Jedes der 12 Items beschreibt einen spezifischen Problembereich von Patienten mit psychischen Störungen, z.B. Item 7 „Gedrückte Stimmung“. Dem Beurteilungsformat liegt eine 5-stufige Likert-Skala von 0 = kein Problem bis 4 = sehr schweres Problem zugrunde. Die 12 Items können auch zu vier Skalen „Verhalten“, „Impairment“, „Symptome“ und „soziale Funktionsfähigkeit“ zusammengefasst werden (Andreas et al., 2007). Darüber hinaus ist es möglich einen Gesamtscore über alle 12 Items zu bestimmen. Hierzu sollten mindestens 10 der 12 Items beantwortet sein.

Tabelle 2: Die vier Dimensionen der HoNOS-D mit den dazugehörigen Items

Dimensionen der HoNOS-D	Items	Itembezeichnung der Dimensionen (Kurzbezeichnung)
Verhalten	Item 1	Überaktives, aggressives, Unruhe stiftendes Verhalten (Aggression)
	Item 2	Absichtliche Selbstverletzung (Selbstverletzung)
	Item 3	Problematischer Alkohol- oder Drogenkonsum (Substanzmissbrauch)
Impairment	Item 4	Kognitive Probleme
	Item 5	Probleme in Zusammenhang mit körperlicher Erkrankung oder Behinderung (Probleme mit körperlichen Erkrankungen)
Symptome	Item 6	Probleme in Zusammenhang mit Halluzinationen oder Wahnvorstellungen (Halluzinationen und Wahnvorstellungen)
	Item 7	Gedrückte Stimmung
	Item 8	Andere psychische und verhaltensbezogene Probleme (Andere Symptome)
Soziale Funktionsfähigkeit	Item 9	Probleme mit Beziehungen
	Item 10	Probleme mit alltäglichen Aktivitäten
	Item 11	Probleme durch die Wohnbedingungen
	Item 12	Probleme durch die Bedingungen im Beruf und im Alltag

Zur Konkretisierung der Schweregradeinschätzungen gibt ein umfangreiches Glossar detaillierte Anweisungen (Anhang 2). Um die differenzierten Operationalisierungen zu demonstrieren, ist in Abbildung 1 ein Beispielitem (Item 7 „Gedrückte Stimmung“) aus dem Glossar aufgeführt. So wird z.B. bei Item „Gedrückte Stimmung“ der Rater angewiesen, eine „1“ (klinisch unbedeutendes Problem ohne Handlungsbedarf) zu vergeben, wenn bedrückte oder geringe Veränderungen in der Stimmung des Patienten bestehen. Dagegen soll der Rater eine „4“ (schweres bis sehr schweres Problem) angeben, wenn eine schwere oder sehr schwere Depression mit Schuld oder Selbstanklage vorliegt. Des Weiteren wird für jedes Item spezi-

fisch angegeben, welche Verhaltensweisen bei diesem Item nicht mit in die Beurteilung einzubeziehen sind (vgl. Abbildung 1).

Item 7 „Gedrückte Stimmung“

* Beziehen Sie Überaktivität oder Agitiertheit, welche auf der Skala 1 eingeschätzt werden, nicht mit ein.

* Beziehen Sie Suizidgedanken oder –versuche, welche auf der Skala 2 eingeschätzt werden, nicht mit ein.

* Beziehen Sie Wahnvorstellungen oder Halluzinationen, welche auf der Skala 6 eingeschätzt werden, nicht mit ein.

0 Kein Problem in Zusammenhang mit gedrückter Stimmung während des Bewertungszeitraums.

1 Bedrückt; oder geringe Veränderungen in der Stimmung.

2 Leichte(s), jedoch eindeutig vorhandene(s) Depression und Leiden (z. B. Schuldgefühle; Verlust des Selbstwertgefühls).

3 Depression mit unangemessenen Selbstvorwürfen; zwanghaft beschäftigt mit Schuldgefühlen.

4 Schwere oder sehr schwere Depression mit Schuld oder Selbstanklage.

Abbildung 1: Beispielim (Item 7) der HoNOS-D („Health of the Nation Outcome Scales“, deutsche Version) zur Illustration der Operationalisierungen der einzelnen Skalenstufen

4. Psychometrische Eigenschaften der HoNOS: Theoretischer Hintergrund

Inzwischen liegen viele verschiedene empirische internationale Studien zu psychometrischen Eigenschaften der HoNOS vor. Um die im Kapitel 5 darzulegenden Befunde anschließend einordnen zu können, soll nun eine zusammenfassende Darstellung der Ergebnisse dieser Arbeiten erfolgen.

Wichtige Kriterien für den Einsatz eines Instrumentes in der Routineversorgung beziehen sich auf die praktikable Anwendung sowie die Reliabilität und Validität.

4.1. Zur Praktikabilität der HoNOS

Wie verschiedene Studien zur Untersuchung der **Praktikabilität** der HoNOS aufzeigen, hat sich das Instrument international mittlerweile im routinemäßigen Einsatz bewährt. So kann sowohl der zeitliche Aufwand zum Ausfüllen des HoNOS-Ratings (\varnothing 2-3 min) als auch das Vorliegen auswertbarer Ratingbögen, vor allem zu Beginn der Behandlung, als zufrieden stellend bewertet werden (Browne et al., 2000; Wing et al., 1998). Auf Itemebene betrachtet, wiesen in der Studie von Eagar et al. (2005) im stationären Setting die Items 11 und 12 die meisten fehlenden Werte auf (ca. 13%). Danach folgten Item 8, Item 9 (je 5%) sowie Item 10 und Item 3 (je 4%). Im ambulanten Bereich wies Item 8 mit 2% die höchste Anzahl fehlender Werte auf (Eagar et al., 2005).

4.2. Zur Reliabilität der HoNOS

Trauer et al. (1999) und Orrell et al. (1999) konnten für einen Großteil der Items zufrieden stellende **Interraterübereinstimmungen** aufzeigen. In diesen beiden Studien ergaben sich allerdings auch Hinweise darauf, dass die Items 11 und 12, welche allgemeine Lebensbedingungen der Patienten erfassen, eher niedrige Übereinstimmungen aufwiesen ($ICC = .33 - ICC = .47$). Dagegen wiesen in der Studie von Brooks (2000) nur drei Items konsistent befriedigende Interraterübereinstimmungen auf. Dieses Ergebnis könnte möglicherweise im Design der Studie begründet sein. Da im Gegensatz zu den Studien von Trauer et al. (1999), Orrell et al. (1999) und Wing et al. (1998) hier mehrere Rater mehrere Patienten einschätzten und nicht jeweils ein Raterpaar einen Patienten (siehe

Tabelle 3). Pirkis et al. (2005) können in ihrem Review zu psychometrischen Eigenschaften des HoNOS aufzeigen, dass insbesondere die Items 4 „Kognitive Probleme“, 7 „Depressive Stimmung“, 8 „Andere Probleme“, 9 „Probleme mit Beziehungen“, 11 „Probleme durch die

Wohnbedingungen“ und 12 „Probleme im Beruf und in der Freizeit“ hinsichtlich der Interrater-Reliabilität die meisten Schwankungen über die Studien hinweg aufweisen.

Einige wenige Studien (Orrell et al., 1999; Wing et al., 1996) haben neben der Interrater-Reliabilität auch die Test-Retest Reliabilität untersucht (siehe Tabelle 3). Dabei weisen die Items 1 „Aggressivität“, 3 „Substanzmissbrauch“, 7 „Depressive Stimmung“ und 10 „Probleme mit alltäglichen Aktivitäten“ die niedrigsten Koeffizienten auf.

Tabelle 3: Interrater-reliabilität (IRR) and Test-Retest-Reliabilität (T-RTR) der Health of the Nation Outcome Scales (HoNOS) in verschiedenen Studien

	Wing et al. (1996) ¹ (N = 52)	Orrell et al. (1999) ¹ (N = 100)	Orrell et al. (1999) ¹ (N = 100)	Wing et al. (1998) ¹ (N = 100; N = 97)	Trauer et al. (1999) ² (N = 50)	Brooks (2000) ² (N = 20)	
HoNOS Items	T-RTR ³	T-RTR ⁴	IRR ⁴	Nott. IRR ³	Man. IRR ³	IRR ³	IRR ³
AGGR	0.61	0.32	0.60	0.97	0.80	0.73	0.61
SV	0.75	0.45	0.82	0.88	0.92	0.83	0.52
ALK	0.88	-0.03	0.65	0.99	0.61	0.86	0.81
KOG	0.79	0.65	0.55	0.81	0.92	0.41	0.50
KÖRP	0.78	0.52	0.67	0.88	0.89	0.62	0.44
WAHN	0.88	0.45	0.80	0.87	0.92	0.83	0.88
DEP	0.78	0.31	0.61	0.84	0.89	0.79	0.60
AND	0.81	0.41	0.27	0.95	0.52	0.61	0.27
BEZ	0.84	0.43	0.39	0.74	0.78	0.60	0.33
AKT	0.74	0.32	0.58	0.71	0.90	0.68	0.20
WOHN	0.85	0.46	0.33	0.83	0.72	0.47	0.28
BERUF	0.84	k.A.	k.A.	0.49	0.51	0.38	0.37

Anmerkungen: IRR = Interrater-reliabilität; T-RTR = Test-retest Reliabilität ; ¹ = UK; ² = Australien; ³ = Intra-klassenkorrelationskoeffizienten; ⁴ = kappa. Nott. = Nottingham; Man. = Manchester; AGGR = Item 1 “Aggression”, SV = Item 2 “Selbstverletzung”, ALK = Item 3 “Alkohol- und Drogenmissbrauch”, KOG = Item 4 “Kognitive Probleme” KÖRP = Item 5 “Körperliche Probleme”, WAHN = Item 6 “Halluzinationen und Wahn”, DEP = Item 7 “Depressivität”, AND = Item 8 “Andere Probleme”, BEZ = Item 9 “Probleme mit Beziehungen“, AKT = Item 10 “Probleme mit alltäglichen Aktivitäten”, WOHN = Item 11 “Probleme mit den Wohnbedingungen”, BERUF = Item 12 “Probleme in Beruf und Freizeit”

Bezogen auf die **interne Konsistenz** (Cronbach’s Alpha) der **Gesamtskala** ist einschränkend darauf hinzuweisen, dass aufgrund der postulierten Unabhängigkeit der 12 Items nur moderate Koeffizienten zu erwarten sind. So berichten Page et al. (2001) Koeffizienten von $\alpha = .59$. Trauer et al. (1999) dagegen fand eine befriedigende interne Konsistenz mit $\alpha = .75$, was sich weitestgehend mit den Befunden von Stedman et al. (1997) und Eager et al. (2005) deckt ($\alpha =$

.73 bzw. $\alpha = .80$).

4.3. Zur Validität der HoNOS

McClelland et al. (2000) untersuchten mittels Inter-Item-Korrelationen die von den Testautoren postulierte Unabhängigkeit der 12 Items. Dabei konnten sie die Befunde von Wing et al. (1998) replizieren (> 2 Interkorrelationen über $r = .40$). Sie untersuchten weiterhin (McClelland et al., 2000), in wie weit die Einzelitems Varianz am Gesamtscore aufklärten und konnten aufzeigen, dass Item 7 „Depressive Stimmung“, Item 8 „Andere Probleme“ und Item 9 „Probleme mit Beziehungen“ die größte Varianz aufklärten (mit 15%, 19% und 14%).

Trauer et al. (1999) überprüfte die von den Testautoren postulierten 4-**Faktorenstruktur** mittels einer konfirmatorischen Faktorenanalyse und fanden nur moderate Gütekriterien für das Modell. Aufgrund dessen berechneten sie eine explorative und eine konfirmatorische Faktorenanalyse und gelangte mit 5 Subskalen zu einer besseren Passung des Modells (siehe Abbildung 2). Eagar et al. (2005) überprüften mittels konfirmatorischer Faktorenanalysen sowohl die Faktorenstruktur der Testautoren als auch die gewonnene Struktur von Trauer et al. (1999). Auch sie kamen zu dem Ergebnis, dass die 5-Faktorenlösung von Trauer et al. (1999) bessere Gütekriterien aufzeigte als die 4-Faktorenlösung von Wing et al. (1998). Auch McClelland et al. (2000) konnten nicht die Faktorenstruktur von Wing et al. (1998), die inhaltlich festgelegt wurde, replizieren. Dabei bestanden in der Studie von Trauer et al. (1999) jeweils für die Skala „Impairment“ und „Verhalten“ sowie „Verhalten“ und „Soziale Probleme“ Interkorrelationen von mindestens großer Effektstärke ($r > .60$), dagegen für die anderen Skalen Interkorrelationen von höchstens mittlerer Effektstärke ($r \leq .30$). Die internen Konsistenzen für die Subskala „Verhalten“ und „Impairment“ waren in der Studie von Trauer et al. (1999) in einem moderaten Bereich ($\alpha = .41$ und $\alpha = .28$). Für die beiden Subskalen „Soziale Probleme“ und „Depression“ befanden sich die internen Konsistenzen in akzeptablen Bereichen mit $\alpha = .71$ und $\alpha = .62$. Zur besseren Veranschaulichung ist die 5-Faktorenstruktur der HoNOS in Abbildung 2 dargestellt.

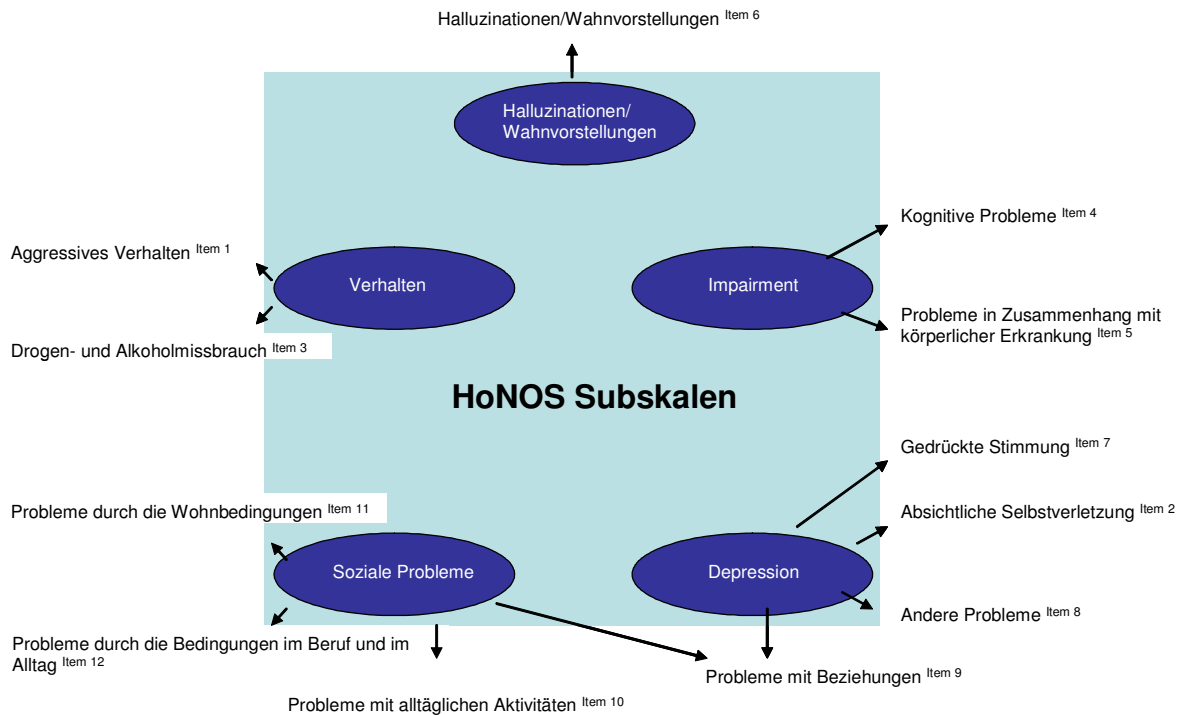


Abbildung 2: Die 5-Faktorenstruktur der HoNOS, modifiziert nach Eagar et al. (2005).

In mehreren Studien wurden die **konkurrente Validität** und die **Änderungssensitivität** der HoNOS untersucht. Dabei zeigte sich, dass der HoNOS weitestgehend in der Lage ist, theoretisch zu erwartende Unterschiede im Alter, Geschlecht und in Diagnosegruppen aufzuzeigen (Bech et al., 2003; Browne et al., 2000; Eagar et al., 2005; McClelland et al., 2000; Trauer et al., 1999). In der Studie von Rees et al. (2004) fanden sich dagegen keine signifikanten Zusammenhänge zwischen der Diagnose und der initialen Schweregradbelastung. Weiterhin konnte in mehreren Studien aufgezeigt werden, dass der HoNOS (Subskalen und Gesamtscore) signifikant mittlere Zusammenhänge zu anderen **fremdeingeschätzten Instrumenten** (z.B. der Global Assessment of Functioning Scale (Endicott et al., 1976)) aufwies (Adams et al., 2000; Browne et al., 2000; Gallagher & Teesson, 2000; Issakidis & Teesson, 1999; McClelland et al., 2000; Orrell et al., 1999; Parker et al., 2002; Salvi et al., 2005).

Bei den Zusammenhängen der HoNOS mit **selbsteingeschätzten Instrumenten** zeigten sich eher heterogenen Ergebnisse. Während eine Vielzahl an Studien signifikante Zusammenhänge zu selbsteingeschätzten Instrumenten (z.B. SCL-90-R oder Short-Form-36) aufzeigten (Adams et al., 2000; Brooks, 2000; Hope et al., 1998; Orrell et al., 1999; Page et al., 2001), konnte Brooks (2000) diese Zusammenhänge nicht finden. Dieses Ergebnis ist u.U. in Abhängigkeit vom zugrunde liegenden Design zu sehen. In der Studie von Brooks (2000) schätzten

mehrere Rater auch mehrere Patienten ein, was wiederum zum größeren Informationsverlust zwischen Rater und Patient oder zu unterschiedlichen Interviewtechniken zwischen den Ratern geführt haben und sich somit nicht nur auf die Reliabilitätswerte (s.o.), sondern entsprechend auch auf die Validitätskennwerte ausgewirkt haben kann.

Ebenso überprüften viele Studien, in wie weit die Einschätzung im HoNOS bei Patienten mit **unterschiedlichen Diagnosegruppen** variiert. Dabei fanden einige Studien heraus, dass höhere Werte im Gesamtscore des HoNOS mit folgenden Diagnosen assoziiert waren: Substanzabhängigkeit und Bipolare Störungen. Höhere Ausprägungen in den Items 1 „Aggressivität“, 3 „Substanzabusus“ und Item 8 „Andere Probleme, Angst“ waren assoziiert mit Manien; dagegen waren höhere Werte in den Items 2 „Suizidalität/Selbstverletzung“, 5 „Körperliche Probleme“ und 7 „Depressive Stimmung“ mit Patienten mit einer Depression (Amin et al., 1999; Bech et al., 2003; Eagar et al., 2005; McClelland et al., 2000; Rees et al., 2004).

Hinsichtlich der **prognostischen Validität** konnten einige Studien aufzeigen, dass der Gesamtscore im HoNOS den Ressourcenverbrauch (z.B. Kontakte, Behandlungsdauer oder Kosten) und das Therapieergebnis (z.B. Wiederaufnahmeraten, Therapieresponse und Sterbefälle) (Broadbent, 2001; Hope et al., 1998; Parker et al., 2002; Schneider et al., 2002) vorhersagen kann. Zwei andere Studien dagegen konnten den Zusammenhang zwischen dem Gesamtscore im HoNOS und im Ressourcenverbrauch nicht aufzeigen (Boot et al., 1997; Goldney et al., 1998).

Auch bezüglich der **Änderungssensitivität** sind die empirischen Befunde heterogen. So stellten mehrere Studien eine signifikante Abnahme der Werte im HoNOS mit mindestens mittlerer Effektstärke nach einer Behandlung fest (Audin et al., 2001; Brooks, 2000; Gallagher & Teesson, 2000; Goldney et al., 1998; Page et al., 2001; Rees et al., 2004; Taylor & Wilkinson, 1997), wohingegen in anderen Studien keine signifikanten Veränderungen im HoNOS aufgezeigt werden konnten (Parabiaghi et al., 2005; Trauer et al., 1999). Kritisch muss dabei zur Studie von Parabiaghi et al. (2005) angemerkt werden, dass die Autoren nicht angaben, zu welchem Zeitpunkt die zweite Einschätzung im HoNOS erfolgte und welche Behandlung durchgeführt wurde. Weiterhin konnte aufgezeigt werden, dass es signifikante Zusammenhänge zwischen der Schwere der Erkrankung (HoNOS Gesamtscore) und dem **Ressourcenverbrauch** von Patienten mit psychischen Störungen gibt (Bech et al., 2003; Broadbent,

2001; Trauer et al., 1999).

Zusammenfassung

Es bestand, wie in Kap. 3 dargelegt, ein **Defizit** an deutschsprachigen Fremdeinschätzungsinstrumenten zur differenzierten Erhebung des Schweregrades von Patienten mit psychischen Störungen, welche auch in der Routineversorgung eingesetzt werden können und das Outcome reliabel und valide erfassen. Im internationalen Bereich hat sich vor allem die Health of the Nation Outcome Scales (HoNOS) bewährt. Es lässt sich aufgrund der aktuellen Befundlage jedoch ableiten, dass die psychometrischen Eigenschaften der HoNOS als heterogen zu bewerten sind. Auch wurde der HoNOS bislang vor allem in Stichproben eingesetzt, die häufig aus Patienten mit Schizophrenen Störungen bestanden. Es liegen derzeit noch vergleichsweise wenige empirische Befunde zu psychometrischen Eigenschaften an Stichproben mit den Diagnosen Affektive Störungen bzw. „Neurotische-, Somatoforme- und Belastungsstörungen“ vor, für die jedoch eine hohe Prävalenz zu verzeichnen ist und die ambulant wie stationär am häufigsten behandelt werden. Auch bezüglich der Ergebnisse zur Validität zeichnet sich ein eher heterogenes Bild ab. Spezifisch betrachtet, betrifft dies vor allem das Item 8 „Andere Probleme“, welches sowohl die meisten fehlenden Werte als auch heterogene Befunde bezüglich der Interrater-Reliabilität aufweist. Daraus lässt sich ein Forschungsbedarf zur psychometrischen Überprüfung der deutschen Version der HoNOS ableiten.

Der damit verbundene Forschungsbedarf geht mit folgenden Fragestellungen einher:

1. Wie viel Aufwand ist nötig, damit der HoNOS gut eingeschätzt werden kann? Wie viele gezielte Fragen sind über ein allgemeines klinisches Interview hinausgehend dazu nötig?
2. Braucht es eine Schulung im HoNOS, damit er von den behandelnden Therapeuten ausgefüllt werden kann?
3. Handelt es sich bei dem HoNOS um eine Sammlung von Einzelskalen oder können Subskalen bzw. ein Gesamtscore verwendet werden?
4. Kann der HoNOS bei unterschiedlichen Patientengruppen und in unterschiedlichen Settings der Versorgung den Schweregrad differenziert messen?
5. Wie sehr schützt eine indirekte Ergebnismessung über wiederholte Beurteilungen im HoNOS-D vor dem Bias der direkten Bewertung der eigenen Arbeit?
6. Kann das Item 8 „Andere Probleme“ die Komorbidität von psychischen Problemen praktikabel, reliabel und valide erfassen?

5. Eigene Studien

5.1. Kontext und Begründung

Instrumentarien zur zuverlässigen und differenzierten Fremdeinschätzung des Schweregrades bei Patienten mit psychischen Störungen gewinnen nicht zuletzt vor dem Hintergrund der zunehmenden Qualitätssicherungsbemühungen seitens der Kostenträger an Bedeutung. In Deutschland gab es bislang kein differenziertes Fremdeinschätzungsinstrument, welches praktikabel in der Routineversorgung und in unterschiedlichen Versorgungssettings einsetzbar gewesen wäre. Mit der Health of the Nation Outcome Scales, HoNOS (Wing et al., 1998) stand international ein weit verbreitetes Instrument zur Verfügung. Die empirischen Befunde zu psychometrischen Eigenschaften der HoNOS erwiesen sich allerdings als heterogen (siehe Kap. 4). Damit war es notwendig, umfangreiche empirische Studien zur Untersuchung der psychometrischen Eigenschaften der deutschen Version der HoNOS (Andreas et al., 2009a; Andreas et al., 2007; Andreas et al., 2010; Andreas et al., in press; Andreas et al., submitted) durchzuführen. Darüber hinausgehend wurde ein für den Versorgungsbereich psychischer Erkrankungen relevantes Item der HoNOS (Item 8 „Andere Probleme), welches u.a. Angst oder Zwanghaftigkeit als komorbides Problem erfasst, weiter entwickelt und bezüglich seiner psychometrischen Eigenschaften untersucht. Ebenfalls wurde die Eignung der HoNOS bei einer spezifischen Patientengruppe, nämlich Patienten mit Abhängigkeitserkrankungen, untersucht.

Die eigenen Untersuchungen wurden im Rahmen eines von der DFG-geförderten Forschungsprojektes „Überprüfung der Validität eines Fremdratings zur differenzierten Erfassung des Schweregrades: „Health of the Nation Outcome Scales“ (GZ AN 382/3-1) und im Rahmen eines vom Norddeutschen Forschungsverbund für Rehabilitationsforschung (NVRF) initiierten Projektes „Entwicklung eines Klassifikationssystems für Patienten mit psychischen Erkrankungen“, welches vom BMBF und der Deutschen Rentenversicherung Bund finanziert wurde (01GD 9808/0), durchgeführt.

5.2. Ziele und Fragestellungen

Aufgrund der aufgeführten empirischen Befunde zu den psychometrischen Eigenschaften der HoNOS haben wir in unserer Arbeitsgruppe von 2004 bis 2009 Studien mit quasi-

experimentellen Design im Versorgungsbereich der stationären psychotherapeutischen Behandlung von Patienten durchgeführt. Die Studien fokussierten die folgenden Fragestellungen:

- a) Als wie praktikabel, reliabel und valide erweist sich die Einschätzung im HoNOS-D?
- b) Ist der HoNOS-D geeignet den Schweregrad differenziert bei Patienten mit Abhängigkeitserkrankungen in stationärer psychotherapeutischer Behandlung zu erfassen?
- c) Als wie praktikabel, reliabel und valide erweist sich die Weiterentwicklung des Items 8 „Andere Probleme“ der HoNOS-D?
- d) Ist der HoNOS-D geeignet den Ressourcenverbrauch von Patienten mit psychischen Erkrankungen in stationärer psychotherapeutischer Behandlung zu erfassen?

5.3. Design und Methodik

Die meisten psychometrischen Untersuchungen der deutschen Version der HoNOS wurden an großen, klinischen und weitest gehend repräsentativen Stichproben von Patienten mit psychischen Erkrankungen in stationärer psychotherapeutischer Behandlung vorgenommen (Andreas et al., 2009a; Andreas et al., 2007; Andreas et al., in press; Andreas et al., submitted).

Die Gesamtstichprobe (N = 3.169 Patienten) wurde zur Untersuchung der Validität der HoNOS herangezogen (Andreas et al., 2010) sowie zur Untersuchung der psychometrischen Eigenschaften des weiterentwickelten Items 8 der HoNOS (Andreas et al., submitted). Eine zeitlich vor der Gesamtstichprobe gewonnene Stichprobe wurde verwendet, um die Praktikabilität und Reliabilität der HoNOS (N = 1.918) sowie die Eignung der HoNOS zur Vorhersage des Ressourcenverbrauchs zu untersuchen (Andreas et al., 2009a; Andreas et al., 2007). Im Folgenden soll die Darstellung des Designs, der Methodik, die Stichprobenbeschreibung, die Beschreibung der verwendeten Instrumente und statistischen Auswertungsverfahren anhand der Gesamtstichprobe erfolgen (**a**).

In Ergänzung zu diesen Analysen wurde vor allem für die Überprüfung von einigen Validitätsaspekten im Rahmen einer separaten Studie (**b**) eine konsekutiv gewonnenen Stichprobe (N = 55) von Patientinnen mit psychischen Erkrankungen in stationärer psychotherapeutischer Rehabilitationsbehandlung untersucht (Andreas et al., 2010).

Zur Überprüfung der spezifischen Eignung der HoNOS bei Patienten mit Abhängigkeitserkrankungen wurde eine multizentrische Studie von Juli 2004 bis Januar 2006 in vier verschiedenen Kliniken zur stationären Rehabilitationsbehandlung Suchtkranker durchgeführt (c) (Andreas et al., in press).

- a) Im Rahmen des Qualitätssicherungsprogramms der Gesetzlichen Krankenkassen (QS-Reha[®]) wurden Daten von insgesamt N = 3.169 konsekutiv aufgenommenen Patienten aus 19 verschiedenen Kliniken des Versorgungsbereichs zu Beginn und zum Ende der Behandlung erhoben. Die Datenerhebungsphase dauerte von Juni 2002 bis Juli 2007. Die Einschätzung im HoNOS wurde durch die behandelnden Therapeuten nach dem klinischen Aufnahme- oder Erstgespräch vorgenommen. Aufgrund von empirischen Befunden (Rock & Preston, 2001) zur Untersuchung der Wirksamkeit eines Trainings im HoNOS wurde kein spezifisches Training der Therapeuten im HoNOS durchgeführt. Die Therapeuten erhielten dennoch ein speziell für die Anwendung der HoNOS in der Routineversorgung entwickeltes Informationsmaterial (Andreas, 2005).
- b) Auf der Basis von Videoaufzeichnungen strukturierter-standardisierter klinischer Interviews (N = 55) von Patientinnen einer psychosomatischen/psychotherapeutischen Rehabilitationsklinik wurden HoNOS Einschätzungen von drei zuvor trainierten und voneinander unabhängigen Ratern durchgeführt. Die Videoaufzeichnungen entstanden in der ersten Woche des stationären Aufenthaltes der Patientinnen. Die standardisierten Interviews wurden zur Diagnostik nach dem Diagnostischen und Statistischen Manual für psychische Störungen, DSM-IV (APA, 2000) verwendet. Der Diagnostiker war gegenüber den Einschätzungen im HoNOS-D blind.
- c) Auch hier wurden die Patienten (N = 417) der 4 verschiedenen Kliniken zur stationären Behandlung Suchtkranker konsekutiv in die Studie aufgenommen. Jeweils ein Raterpaar (insgesamt 16 Therapeuten) schätzten den Schweregrad der Patienten im HoNOS zu Beginn der Behandlung ein. Auch hier wurden die Einschätzungen nach dem klinischen Aufnahme- bzw. Erstgespräch durchgeführt. Nach ca. 60 Ratings pro Klinik wurde ein Training der Therapeuten vorgenommen. Die Schulungsmaterialien wurden hierbei in Anlehnung an Wing et al. (1996) und Trauer (2003) in die deutsche Sprache übersetzt und überarbeitet. Sowohl die Einschätzungen vor als auch nach dem

Training wurden für die Auswertungen verwendet. Zusätzlich zu den Einschätzungen im HoNOS wurden soziodemographische (u.a. Alter, Geschlecht, Schulabschluss) und weitere klinische Parameter (u.a. Diagnosen) erhoben. Am Ende der Behandlung wurde zur Überprüfung der prognostischen Validität der HoNOS Einschätzungen in der Klassifikation Therapeutischer Leistungen, KTL (Bundesversicherungsanstalt für Angestellte (BfA), 2000) dokumentiert.

5.4. Stichprobenbeschreibung

- a) Es wurden insgesamt 3.743 Patienten in die Studie aufgenommen. Davon lagen, wie oben bereits erwähnt, 3.169 vollständige Datensätze sowohl zu Beginn als auch zum Entlasszeitpunkt vor (siehe Tabelle 4). N = 301 Patienten nahmen aus verschiedenen Gründen (z.B. Verweigerung) gar nicht an der Studie teil, von 273 Patienten lagen nur unvollständige Datensätze vor. Drop-out Analysen zwischen den Studienteilnehmern und den Nichtteilnehmern ergaben in relevanten klinischen Parametern (z.B. Diagnose oder Chronifizierung der Erkrankung) eine geringe Anzahl an signifikanten Unterschieden mit mindestens kleiner Effektstärke. Damit konnte die Stichprobe der Studienteilnehmer als weitest gehend repräsentativ bewertet werden. Neben den soziodemographischen und klinischen Merkmalen, die der Tabelle 4 zu entnehmen sind, war die durchschnittliche Behandlungsdauer bei $M = 43$ Tagen ($SD = 17$ Tage, Range: 1-146 Tage).
- b) Initial wurden 65 Patientinnen um Teilnahme an der Studie gebeten. Davon lehnten 10 Patientinnen aus verschiedenen Gründen eine Teilnahme ab. Auch hier bestand kein signifikanter Unterschied zwischen den Studienteilnehmerinnen und den Drop-out Patientinnen bezüglich soziodemographischer oder klinischer Merkmale. Die soziodemographischen und klinischen Merkmale sind ebenfalls der zu entnehmen. Darüber hinausgehend hatten fast alle Patientinnen (95%) vor der stationären Behandlung an einer ambulanten Therapie teilgenommen. Die durchschnittliche Chronifizierung betrug 10 Jahre ($SD = 6$ Jahre). Die durchschnittliche Symptombeeinträchtigung zu Beginn der Therapie mit $M = 1.29$ ($SD = 0.62$) lag in einem für diesen Versorgungsbereich repräsentativen Bereich. Die durchschnittliche Dauer der Behandlung betrug 78 Tage ($SD = 26.09$).

- c) Die Stichprobe für die Untersuchung der spezifischen Eignung der HoNOS-D bestand aus insgesamt N = 417 Patienten mit Abhängigkeitserkrankungen. Zum Zeitpunkt der Aufnahme in die stationäre Rehabilitationsbehandlung hatten alle Patienten eine Entzugsbehandlung absolviert und waren für mindestens 2 Wochen abstinent. Die soziodemographischen und klinischen Merkmale der Stichprobe sind ebenfalls in Tabelle 4 dargestellt.

5.5. Verwendete Instrumente

In den hier zusammen gefassten Studien kamen zur Überprüfung der psychometrischen Kennwerte der HoNOS-D standardisierte und weit verbreitete Instrumentarien zur Selbst- und Fremdeinschätzung zur Anwendung.

Selbsteinschätzungsverfahren

- *Lang- und Kurzversion der Symptom Checklist* SCL-90-R und SCL-14 (Derogatis, 1977; Franke, 1995; Harfst et al., 2002; Prinz et al., 2008): In der Langversion 90 Items und in einer Kurzversion 14 Items enthaltene Symptom Checkliste zur Erfassung der Symptomschwere von Patienten mit psychischen Erkrankungen. Die Patienten beurteilen ihre subjektiv erlebten Beeinträchtigungen durch verschiedene Symptome auf einer 5-stufigen Skala von „überhaupt nicht“ (0) bis „sehr stark“ (4). Die abgefragten Symptome werden zu den folgenden 3 Skalen zusammengefasst: „Depressivität“, „Phobische Angst“ und „Somatisierung“. Die internen Konsistenzen der drei Subskalen lassen sich mit Werten zwischen $r = 0,80$ (Somatisierung) und $r = 0,87$ (Depressivität) als zufrieden stellend bewerten. Der Wert für den Gesamtscore (GSI-14) liegt mit $r = 0,87$ im ebenfalls zufrieden stellenden Bereich und die konvergente Validität zwischen dem GSI-14 und dem GSI der SCL-90-R kann mit $r = 0,93$ erwartungsgemäß als hoch bezeichnet werden (Harfst et al., 2002). Eine ausführliche Darstellung der psychometrischen Eigenschaften der SCL-90R und ihrer Kurzversionen findet sich bei Prinz et al. (2008).
- *Kurzversion des Inventars zur Erfassung interpersoneller Probleme* IIP-25 (Harfst et al., 2004): Harfst et al. (2004) entwickelten an einer Stichprobe von N = 1,683 Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung mittels exploratorischer und konfirmatorischer Faktorenanalysen eine Kurzversion des

Inventars zur Erfassung Interpersoneller Probleme (Horowitz et al., 1994). Die internen Konsistenzen der 5-faktoriellen Lösung betragen zwischen $r = 0,76$ und $r = 0,84$ und können als befriedigend bewertet werden. Hinsichtlich der Kriteriumsvalidität zeigten sich ebenfalls größtenteils befriedigende Ergebnisse. So wiesen z.B. Patienten mit einer sozialen Phobie signifikant höhere Werte auf der Subskala des IIP-25 „sozial-vermeidend“ auf.

- *Short-Form 8 SF-8* (Ware et al., 2000): Der SF-8 ist ein generisches Instrument zur Erfassung des Gesundheitsstatus eines Patienten (Ware et al., 2000). Ziel bei der Konstruktion der SF-8 war es, eine Kurzform des SF-36 zu entwickeln, in der 1 Item für die 8 Gesundheitskonzepte der SF-36 steht und welches in größeren Studien praktikabel einsetzbar sein sollte. Das Antwortformat der SF-8 variiert von 5- bis 6-stufiger Skala. Sowohl mit der SF-8 als auch mit dem SF-36 lassen sich zwei Summenskalen („Psychische und Körperliche Summenskala“) bilden. Zur Reliabilität lässt sich festhalten, dass interne Konsistenzen in einem zufrieden stellenden Bereich zwischen .70 und .88 variieren. Eine Überprüfung der Validität wurde in insgesamt 14 Studien durchgeführt: Konvergente Validität ist gegeben (die Skalen der SF-8 korrelieren signifikant mit den Skalen des SF-36) und unabhängige Faktorenanalysen anhand großer repräsentativer amerikanischer Stichproben konnten die 2-Faktorenlösung der SF-8 replizieren. In klinischen Studien (Personen ohne bzw. mit körperlichen oder psychischen Problemen) konnte die diskriminative Validität der SF-8 gut belegt werden: Die körperliche Summenskala wies in der Gruppe der Personen mit körperlichen Problemen die höchsten Werte auf, die Psychische Summenskala dagegen in der Gruppe der Personen mit psychischen Problemen. Auch für die SF-8 ist eine Normierung im Rahmen einer amerikanischen Gesundheitsuntersuchung durchgeführt wurden.

Tabelle 4: Stichprobencharakteristika der Stichproben a), b) und c)

	<i>a)</i> <i>Andreas et al.,</i> <i>2010</i>		<i>b)</i> <i>Andreas et al.,</i> <i>2010</i>		<i>c)</i> <i>Andreas et al.,</i> <i>in press</i>	
	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
Geschlecht						
Weiblich	2290	72	55	100	63	17
Männlich	871	28	0	0	314	83
Alter (M = 45, SD = 14)			(M = 31, SD = 9)		(M = 34, SD = 10)	
< 29 Jahre	469	15	32	58	133	37
30 - 39 Jahre	565	18	14	25	113	31
40 - 49 Jahre	896	28	7	13	80	22
50 - 59 Jahre	766	24	2	4	23	6
> 60 Jahre	462	15	0	0	14	4
Familienstand						
verheiratet	1387	44	8	15	66	16
ledig	900	29	40	73	234	56
geschieden	430	14	3	6	45	11
in Trennung	221	7	9	16	10	2
verwitwet	199	6	0	0	6	1
registrierte Partnerschaft	<i>n.e.</i>		<i>n.e.</i>		7	2
Höchster Schulabschluss						
Realschulabschluss	1048	34	11	20	70	23
Hauptschulabschluss	981	32	5	9	158	53
Abitur	821	27	39	71	26	9
noch in der Schule, sonstiger Schulabschluss	179	6	0	0	1	0
kein Schulabschluss, Sonderschule	60	2	0	0	46	15
Höchster beruflicher Abschluss						
Lehre/Fachschule	1381	50	<i>n.e.</i>	<i>n.e.</i>	64	21
Fachhochschule/Uni	535	20	<i>n.e.</i>	<i>n.e.</i>	5	2
ohne Berufsabschluss	324	12	<i>n.e.</i>	<i>n.e.</i>	4	1
noch in Berufsausbildung	104	4	<i>n.e.</i>	<i>n.e.</i>		
Meister	122	4	<i>n.e.</i>	<i>n.e.</i>		
sonstiges	283	10	<i>n.e.</i>	<i>n.e.</i>		
Erwerbstätigkeit						
berufstätig	1558	51	22	40	39	13
Hausfrau/-mann	334	13	15	27	2	1
arbeitslos	385	11	<i>n.e.</i>	<i>n.e.</i>	65	22
Altersrente	97	6	<i>n.e.</i>	<i>n.e.</i>	11	1
EU -Rente	166	3	<i>n.e.</i>	<i>n.e.</i>		
sonstiges	501	17	16	29	108	38

Tabelle 4 fortgesetzt:

Diagnosen	(ICD-10; Erstdiagnose)		(DSM-IV; Komorbiditäten)		(ICD-10; Erstdiagnose)	
Psychische und Verhaltensstörungen durch...						
F10: Alkohol					106	31
F11: Opioide					74	21
F12: Cannabinoide					42	12
F13: Sedativa oder Hypnotika					2	1
F14: Kokain					29	8
F15: andere Stimulanzien, einschließlich Koffein					5	1
F19: multiplen Substanzgebrauch und Konsum anderer psychotroper Substanzen					58	17
Depression (F32-33, F34.1)	1077	35	35	64	1	0
Anpassungsstörung/Neurasthenie (F43.2-9, F48)	486	16	18	33		
Essstörung	320	10	26	47		
Angststörung (F40-41)	288	9	20	36		
somatoforme Störung (F45)	190	6	13	24		
Persönlichkeitsstörung (F60)	114	4	46	84		
Schizophrenie, schizotype und wahnhaftige Störung	97	3	0	0		
sonstige Diagnosen	333	11	25	46	28	7

Anmerkungen: n.e. = nicht erhoben

- *Beck Depressionsinventar* BDI (Beck et al., 1961; Hautzinger et al., 1995): 21-Item umfassendes Selbsteinschätzungsinstrument zur Erfassung der Schwere depressiver Symptomatik. Items beziehen sich auf schwere depressive Symptomatik wie z.B. persönliche Schwächen, Schuldgefühle oder sozialer Rückzug, die auf einer 4-stufigen Likert-Skala beantwortet werden. Interne Konsistenzen können als gut bis sehr gut bewertet werden (Cronbach's $\alpha = 0.74 - 0.92$). Ebenso zeigen sich zufrieden stellende Zusammenhänge mit anderen Selbsteinschätzungsinstrumenten, die Depressivität erfassen ($r = 0.72 - 0.74$) (Hautzinger et al., 1995).

Fremdeinschätzungsverfahren

- *Short-Form 8* SF-8 (Schulz et al., 2009; Ware et al., 2000) ist die Fremdeinschätzungsversion der SF-8 (siehe Beschreibung oben).

- *Global Assessment of Functioning Scale* (Endicott et al., 1976) (ausführliche Beschreibung siehe Kap. 2)
- *Beeinträchtigungs-Schwere-Score*, BSS, (Schepank, 1995)(ausführliche Beschreibung siehe Kap. 2)

Zur spezifischen Überprüfung der Eignung der HoNOS-D zur Vorhersage des Ressourcenverbrauchs wurden Angaben zu verbrauchten therapeutischen Leistungen gesammelt.

- *Kategorien Therapeutischer Leistungen (KTL)*: Dokumentation therapeutischer Leistungen am Ende einer stationären psychosomatischen/psychotherapeutischen Rehabilitationsbehandlung (Bundesversicherungsanstalt für Angestellte (BfA), 2000).

5.6. Statistische Analysen

Es wurden psychometrische Analysen zur Überprüfung der Praktikabilität, Reliabilität und Validität der HoNOS-D vorgenommen.

(1) Statistische Auswertungsverfahren zur Untersuchung der Praktikabilität und Reliabilität

Aus:

Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology, 40*, 116-125.

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research, 19*, 50-62.

Aufgrund der empirischen Befunde aus der Literatur wurde erwartet, dass die meisten Items der HoNOS-D auch in unserer Studie zufrieden stellende Werte bezüglich der **Praktikabilität** aufweisen sollten (Rate fehlender Werte < 10%). Für 3 der 12 Items wurden höhere Raten an fehlenden Werten erwartet: Für Item 8 „Andere Probleme“, Item 11 „Probleme mit den

Wohnbedingungen“ und Item 12 „Probleme in Beruf und Freizeit“. Zur Itemverteilung kann auf der Basis der Ergebnisse aus der Literatur angenommen werden, dass Item 3 „Substanzmissbrauch“ und Item 6 „Halluzinationen und Wahnvorstellungen“ eher schief verteilt sind (Audin et al., 2001). Neben einer Analyse fehlender Werte, wurden auch deskriptive Itemanalysen durchgeführt.

Vor dem Hintergrund der spezifischen Testkonstruktion der HoNOS wurden niedrige Korrelationen zwischen den Items erwartet ($r < 0,30$) (Wing et al., 1998). Dazu wurde eine **Iteminterkorrelationsmatrix** (Spearman-Rangkorrelationen) berechnet.

Basierend auf den Ergebnissen zur **Test-Retest-Reliabilität** aus anderen Studien wurden höhere Korrelationen zwischen der ersten und der zweiten Messung für die meisten der HoNOS-D Items erwartet (Wing et al., 1998). Dazu wurden Intraklassenkorrelationskoeffizienten (ICC) zwischen dem Beginn und dem Ende der Behandlung bestimmt (Everitt, 1996). Aufgrund dessen, dass zwischen den beiden Messzeitpunkten eine Behandlung stattgefunden hatte, wurde die Test-Retest-Reliabilität nur für diejenigen Patienten berechnet, die sich hinsichtlich ihres Therapieergebnisses nicht statistisch signifikant verändert hatten (“no significant change group”, $n = 539$). Zur Bestimmung der statistisch signifikanten Veränderungen wurde der Reliable Change Index (RCI) von Jacobson (1991) zugrunde gelegt. Dieser bildete sich aus der Psychischen Summenskala des SF-8 (Fremd- und Selbsteinschätzung) unter Verwendung der folgenden Formel:

$$RCI = 2 * \sqrt{2} * SD_1 * \sqrt{(1 - REL)} \quad (\text{Jacobson \& Truax, 1991})$$

(SD_1 = Standardabweichung der Patientenpopulation zu t_1 , REL = Test-retest Reliabilität des SF-8, (Ware et al., 2000))

Daraus ergaben sich folgende Werte:

$$RCI_{SF-8-F} = 2 * \sqrt{2} * 0.648 * \sqrt{(1 - 0.82)} = .78$$

$$RCI_{SF-8-S} = 2 * \sqrt{2} * 0.892 * \sqrt{(1 - 0.82)} = 1.07$$

Die Subgruppe (“no significant change group”) wurde dementsprechend wie folgt definiert:

No significant change group = $(RCI_{SF-8-F} \leq 0.78) \& (RCI_{SF-8-S} \leq 1.07)$

In der Arbeit von Andreas et al. (in press) wurden Überprüfungen zur **Interraterreliabilität** der Items der HoNOS-D vorgenommen. Zur Bewertung der Interraterreliabilitäten wurden die Kriterien von Fleiss und Cohen (Fleiss & Cohen, 1973) zugrunde gelegt, die zeigen konnten, dass Interraterreliabilitätskoeffizienten (ICC) zu gewichteten Kappakoeffizienten äquivalent sind. Landis und Koch (1977) konnten „rules of thumb“ für gewichtete Kappakoeffizienten bereit stellen: Kappawerte zwischen 0,21 and 0,40 sind als ‘fair’, zwischen 0,41 and 0,60 als ‘moderat’, zwischen 0,61 and 0,80 als ‘zufrieden stellend’, und zwischen 0,81 and 1,00 als ‘fast perfekt’ anzusehen.

(2) Überprüfungen zur Validität der HoNOS-D

Aus:

Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

In zwei Arbeiten (Andreas et al., 2010; Andreas et al., in press) wurde die **faktorielle Validität** der HoNOS-D an jeweils unterschiedlichen Stichproben (vgl. Kap. 4 Stichprobe a) und c)) untersucht. Dazu wurde die post-hoc postulierte Skalenstruktur der Testautoren (Wing et al., 1998) mittels konfirmatorischer Faktorenanalysen (KFA) überprüft. Da die meisten Items der HoNOS-D eine schiefwinkelige Verteilung aufwiesen, wurde die KFA auf der Basis polychorischer Korrelationsmatrixen mittels des Programms Structural Equations Program, EQS (Bentler & Wu, 2004) vorgenommen. Wie im theoretischen Hintergrund dieser Arbeit bereits dargelegt, überprüften Trauer (1999) und Lauzon et al. (2001) ebenfalls mittels KFA die Faktorenstruktur der HoNOS. Zur Bewertung der Fitindizes und Modellgüte wurde die eher als konservativ zu bewertenden Kriterien von Hu und Bentler (1999) zugrunde gelegt:

- Der Standardized-Root-Mean-Square-Residual SRMR (Bentler, 1990): $SRMR \leq 0.08$ (or 0.10)
- Root-Mean-Square-Error-of-Approximation RMSEA (Steiger, 1990): $RMSEA \leq 0.06$
- Comparative Fit Index CFI (Bentler, 1990): $CFI \sim 0.95$.

Des Weiteren wurden **konvergente** und **diskriminative Validität** mittels Spearman-Rangkorrelationen zwischen den Items bzw. dem Gesamtscore der HoNOS-D und den ent-

sprechenden Skalen der anderen Instrumente untersucht (Andreas et al., 2010). Hierbei wurde erwartet, dass die Items bzw. der Gesamtscore der HoNOS-D mit anderen konvergenten Skalen von selbst eingeschätzten Instrumenten mit mindestens einer mittleren Effektstärke nach Cohen (1988) korrelieren (z.B. Item 7 „Gedrückte Stimmung“ mit der Depressivitätsskala der SCL oder Item 9 „Probleme in Beziehungen“ mit dem Gesamtscore des IIP). Für die Zusammenhänge zwischen den Items bzw. dem Gesamtscore der HoNOS-D zu anderen konvergenten Skalen fremd eingeschätzter Instrumente wurden signifikante Korrelationen mit mindestens großer Effektstärke nach Cohen (1988) ($r > 0,50$) erwartet (z.B. Item 5 „Körperliche Probleme“ mit der Körperlichen Summenskala der SF-8-F). Geringere Korrelationen ($r < 0,10$) wurden für Zusammenhänge zu divergenten Skalen erwartet (z.B. Item 1 „Überaktivität und Aggressivität“ mit der Körperlichen Summenskala der SF-8-F). Um die Hypothesen zu prüfen, wurden durchschnittliche Fishers-Z Korrelationen über alle Einzelkorrelationen berechnet, welche zur Bewertung der divergenten und konvergenten Validität herangezogen wurden.

Zur Überprüfung der **Kriteriumsvalidität** wurde die Stichprobe b) (siehe Kap. 5) herangezogen. Dazu wurde untersucht, in wie weit sich einzelne HoNOS-D Items bzw. der Gesamtscores zwischen unterschiedlichen diagnostischen Gruppen nach DSM-IV (American Psychiatric Association, 1996) bezüglich der Schweregradausprägung unterscheiden. Die Berechnungen wurden mittels Mann-Whitney U Test für Daten auf Rangskalenniveau vorgenommen. Um eine mögliche Kumulierung des alpha-Fehlerwahrscheinlichkeitsniveaus durch Mehrfachtestung entgegenzuwirken, wurde das alpha-Fehlerwahrscheinlichkeitsniveau nach der Bonferroni-Holm Methode korrigiert (Holm, 1979). Die diagnostischen Gruppen wurden dabei wie folgt gebildet:

- Affektive Störungen (DSM-IV: 296.20, 296.3, 300.4)
- Substanzstörungen (DSM-IV: 303.00 - 305.90)
- Borderline Persönlichkeitsstörung (DSM-IV: 301.83) und
- Somatoforme Störungen (DSM-IV: 300.11, 300.81 - 300.82, 307.42 - 307.44, 307.80).

Zur Bewertung der signifikanten Unterschiede wurde Cohen's d (1988) zugrunde gelegt (0,20 $< d < 0,50$ geringe Effektstärke –ES–, 0,50 $< d < 0,80$, mittlere ES und $d > 0,80$, große ES). Auf der Basis empirischer Ergebnisse früherer Studien wurde erwartet, dass Patienten mit spezifischen Diagnosen (z.B. Affektive Störungen) einen höheren Score im Indexitem (z.B.

Item 7 „Gedrückte Stimmung“) der HoNOS-D als Patienten ohne diese Diagnose aufweisen sollten.

Zur Untersuchung der **Änderungssensitivität** der HoNOS-D wurden prä-post Vergleiche mittels Wilcoxon-Test für abhängige Stichproben durchgeführt. Basierend auf der Formel von Kazis et al. (1989) wurde eine standardisierte prä-post ES kalkuliert: $(M1 - M2) / SD1$ ($M1$ = Mittelwert zum Beginn der Behandlung, $M2$ = Mittelwert zum Ende der Behandlung, $SD1$ = Standardabweichung zum Beginn der Behandlung). Effektstärken (ES) wurden auch hier nach den Kriterien von Cohen (1988) interpretiert.

Zur Untersuchung der **prognostischen Validität** und spezifischen Eignung der HoNOS-D in Bezug zum Ressourcenverbrauch in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung und Rehabilitationsbehandlung Suchtkranker kam ein Entscheidungsbaumverfahren zur Anwendung (Andreas et al., 2009a; Andreas et al., 2010). Dabei wurde erwartet, dass Einzelitems der HoNOS-D bzw. der Gesamtscore signifikante Prädiktoren des Ressourcenverbrauches sind. Zur ausführlichen Beschreibung der Methode der Entscheidungsbaumverfahren oder Regression-Tree (siehe Anhang 5 und Anhang 6).

(3) Statistische Analysen zur Überprüfung des weiterentwickelten Items 8 der HoNOS-D

Aus: Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology*, 40, 116-125.

Das differenziertere Rating des weiterentwickelten Items 8 der HoNOS-D sollte zu einer Erhöhung der Präzision in der Beurteilung führen, weswegen wir höhere **Schweregrade** für das Item im Vergleich zur Originalversion mit mindestens geringen ES erwarteten. Diese Hypothese wurde mittels ANOVA überprüft. Die Bewertung der ES folgte wieder den Empfehlungen von Cohen (1988): $\text{partial } \eta^2 > 0.0099$ (geringer Effekt), $\text{partial } \eta^2 > 0.0588$ (mittlerer ES), und $\text{partial } \eta^2 > 0.1379$ (große ES). Um Kohorteneffekte zwischen den verschiedenen zugrunde gelegten Stichproben (eine Stichprobe aus der Pilotphase und eine Stichprobe aus der Routinephase des QS-Reha[®]-Projektes) zu kontrollieren, wurde als Kontrollvariable die durchschnittliche Bewertung der anderen Items der HoNOS und die Ausgangswerte der Körperlichen und Psychischen Summenskala des SF-8 mit in die Berechnungen aufgenommen.

Das weiterentwickelte Item 8, welches zu einer differenzierteren Beurteilung führen sollte, sollte auch im Vergleich zur Item 8 Originalversion mit einer Abnahme an fehlenden Werten und damit mit einer Erhöhung der **Praktikabilität** einhergehen. Das wurde mittels χ^2 -Test geprüft. Als Effektstärkenmaß wurde phi zugrunde gelegt: $\phi > 0.10$ (geringe ES), $\phi > 0.30$ (mittlere ES), und $\phi > 0.50$ (große ES) (Cohen, 1988).

Zur Prüfung der **Interraterreliabilität** des weiterentwickelten Items 8 der HoNOS-D wurden ICCs zwischen drei Ratern, die den HoNOS-D zu Beginn der Behandlung beurteilten, berechnet. Die Berechnungen erfolgten auf der Basis von Stichprobe b) (siehe Kap. 5).

Die **konvergente** und **diskriminante Validität** wurde mittels Person-Produkt-Moment-Korrelationen zwischen den Problembereichen des Items 8 und den selbst eingeschätzten Subskalen der SCL und des SF-8 berechnet. Hierbei wurde erwartet, dass die symptom-spezifischen Probleme signifikant mit mindestens mittlerer ES ($r > 0,30$) nach Cohen (1988) korrelieren sollten (z.B. „Phobisch“ des Item 8 mit der „Phobischen Subskala“ der SCL-14). Signifikante Korrelationen mit fremd eingeschätzten Skalen der SF-8 mit mindestens großen ES ($r > 0,50$) nach Cohen (1988) wurden erwartet (z.B. „Somatoform“ des Item 8 mit der „Körperlichen Summenskala“ des SF-8-F).

Auch hier wurde die **Kriteriumsvalidität** mit dem oben bereits beschriebenen Vorgehen geprüft. Für diese psychometrische Überprüfung wurden allerdings folgende Diagnosegruppen gebildet:

- Angststörungen (DSM-IV: 300.00x, 300.29, 300.01, 300.23, 300.01-300.22)
- Essstörungen (DSM-IV: 307.1 - 307.51)
- Posttraumatische Belastungsstörung (PTSD, DSM-IV: 309.81) und
- Somatoforme Störungen (DSM-IV: 300.81).

Es wurde hierbei erwartet, dass u.a. Patienten mit Angststörungen höhere Schweregradausprägungen im Bereich „Angst“ des weiterentwickelten Item 8 der HoNOS-D aufweisen als Patienten ohne Angststörungen.

Auch in dieser Studie wurde eine Überprüfung der Änderungssensitivität des weiterentwickelten Items 8 im Vergleich zum Item 8 in der Originalversion vorgenommen. Das Vorgehen wurde im vorherigen Abschnitt bereits beschrieben.

6. Ergebnisse

Im Folgenden werden die Ergebnisse der publizierten Studien, die im Anhang 3 bis 7 zu finden sind, kurz zusammenfassend beschrieben. Die Darstellung wird sich auf die für diesen Kontext relevantesten Resultate beschränken. Es werden die empirischen Befunde zu den psychometrischen Eigenschaften **Praktikabilität**, **Reliabilität** und **Validität** der HoNOS-D und des weiter entwickelten Items 8 zusammenfassend für alle untersuchten Stichproben referiert.

6.1. Praktikabilität

Aus:

Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology, 40*, 116-125.

Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.

Zur Überprüfung der **Praktikabilität** der HoNOS-D wurde eine Analyse fehlender Werte durchgeführt. Diese ergab weitest gehend zufrieden stellende Ergebnisse für 11 der 12 HoNOS-D Items mit einer durchschnittlichen Rate von 3,3% (Range: 1,3% bis 4,5%). Zum Beispiel wies das Item 7 „Gedrückte Stimmung“ eine Rate von 1,3% fehlender Werte auf. Wie bereits erwartet, zeigte sich in dieser Analyse, dass das Item 8 „Andere psychische und Verhaltensprobleme“ die höchste Rate mit 8,6% fehlender Werte aufwies.

Damit erwies sich die HoNOS-D als ein weitest gehend praktikables Instrument, welches in der Routineversorgung eingesetzt werden könnte. Es zeigte sich auch, dass **Item 8 „Andere psychischen und verhaltensbezogene Probleme“** nur moderate Resultate aufwies. Betrachtet man die Konzeption des Items 8 und seine Beschreibung im Glossar, so fällt auf, dass dieses Item keine differenzierten Ankerbeispiele enthält. Des Weiteren soll der Beurteiler aus einer Liste von 10 Problemen, die alle unter Item 8 subsumiert sind (z.B. Angst oder Essen), das bedeutsamste Problem herausuchen und hinsichtlich der Schwere der Problematik einschätzen. Damit war weiterer Forschungs- und Entwicklungsbedarf verbunden.

Es schloss sich eine eigene Studie an, die die **Weitentwicklung des Items 8 „Andere psychische und verhaltensbezogene Probleme“** beinhaltet (Andreas et al., submitted). Um eine Vergleichbarkeit und damit Überprüfbarkeit zur Originalversion des Items 8 zu gewährleisten, wurde der Beurteiler in der weiterentwickelten Version darum gebeten, eine differenzierte Beurteilung jedes der 10 Probleme des Items 8 durchzuführen und im Anschluss das schwerwiegendste Problem raussuchen sowie es mit einem Buchstaben und einer Schweregradeinschätzung zu versehen (siehe Anhang 1).

Da in der Pilotphase des QS-Reha[®]-Projektes die Originalversion der HoNOS-D und in der Routinephase die im Item 8 weiterentwickelte Version zum Einsatz kamen, konnte ein direkter empirischer Vergleich zwischen den Beurteilungen in der Pilot- und in der Routinephase durchgeführt werden.

Bezüglich der **Praktikabilität** zeigte sich, dass sich die Anzahl fehlender Werte signifikant mit mindestens kleiner Effektstärke reduzierte (Weiterentwicklung: Auswahl des Problems 8% vs. Originalversion 12% sowie Einschätzung des Schweregrades Weiterentwicklung: 5% vs. Originalversion 8%). Die weiterentwickelte Version des Items 8 der HoNOS-D führte ebenfalls, wie erwartet, zu einer differenzierten Beurteilung. Dieses zeigte sich in signifikant höheren Schweregraden für 7 der 10 Probleme, die unter Item 8 subsumiert sind (z.B. für den Bereich „Sexualität“ mit $\eta^2 = 0.13$). Keine signifikanten Unterschiede zwischen der Weiterentwicklung und der Originalversion wurden für die Bereiche „Phobisch“, „Dissoziativ“ und „Anderes“ gefunden.

6.2. Itemkennwerte der HoNOS-D

Aus:

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.

Die **Verteilungseigenschaften** der Items der HoNOS-D wurden sowohl an der Stichprobe von Patienten mit psychischen Erkrankungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung als auch an der Stichprobe von Patienten mit Abhängigkeitserkrankungen untersucht (siehe Tabelle 4, Kap. 5).

In der Stichprobe von Patienten mit psychischen Erkrankungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung wurden geringere durchschnittliche Ausprägungen ($M < 1,00$) für 5 der 12 HoNOS-D Items gefunden (z.B. Item 1 „Aggressivität“), wohingegen Item 7 „Depressivität“, Item 8 „Andere Probleme“ und Item 9 „Probleme mit Beziehungen“ die höchsten durchschnittlichen Schweregrade aufwiesen (siehe Tabelle 5). Bei den Patienten mit Abhängigkeitserkrankungen zeigte sich erwartungsgemäß die höchsten Ausprägungen im Schweregrad für das Item 3 „Alkohol- und Drogenmissbrauch“, gefolgt von Item 9 „Probleme mit Beziehungen“ und Item 12 „Probleme mit den Wohnbedingungen und Aktivitäten“.

Für die 12 HoNOS Items zeigten sich unterschiedliche Itemverteilungen. Während Item 7 „Depressivität“, Item 8 „Andere Probleme“, Item 9 „Probleme mit Beziehungen“ und Item 10 „Probleme mit alltäglichen Aktivitäten“ zufrieden stellende Verteilungseigenschaften erkennen ließen, zeigten andere Items, insbesondere Item 6 „Halluzinationen und Wahnvorstellungen“ Deckeneffekte. Dieses Ergebnis kann damit zusammenhängen, dass in der Stichprobe auch keine Patienten mit Diagnosen, die diese spezifischen Probleme aufweisen könnten, vorhanden waren (z.B. Patienten mit einer F2 Diagnose nach ICD-10) (World Health Organisation, 1992). Des Weiteren stehen diese Ergebnisse in Einklang mit der Studie von Audin et al. (2001), welche in ihrer Studie Patienten in ambulanter psychotherapeutischer Behandlung im HoNOS eingeschätzt haben. Auch Audin et al. (2001) fanden moderate Verteilungseigenschaften für die Mehrzahl der Items der HoNOS.

In der Stichprobe von **Patienten mit Abhängigkeitserkrankungen** wiesen folgende Items zufrieden stellende Verteilungseigenschaften auf: Item 3 „Alkohol- und Drogenmissbrauch“, Item 7 „Depressivität“, Item 8 „Andere Probleme“, Item 9 „Probleme mit Beziehungen“ und Item 12 „Probleme mit Beruf und Freizeit“. Dieses Ergebnis spricht dafür, wenn Patienten mit korrespondierenden Diagnosen (z.B. F1 „Alkoholabhängigkeit“ nach ICD-10) in der Stichprobe vertreten sind, dass die Items auch zufrieden stellende Verteilungseigenschaften aufweisen.

Tabelle 5: Verteilungseigenschaften der Items der HoNOS-D bei Patienten mit psychischen Erkrankungen in stationärer psychosomatischer/psychotischer Rehabilitationsbehandlung (a) und bei Patienten mit Abhängigkeitserkrankungen nach Entzug in stationärer Rehabilitationsbehandlung (c)

HoNOS Items	M_a (SD_a)	M_c (SD_c)	Decken- effekt _a (%)	Decken- effekt _c (%)	Boden- effekt _a (%)	Boden- effekt _c (%)	Schiefe _a	Schiefe _c	Kurtosis _a	Kurtosis _c	r_{it} (a)	r_{it} (c)
AGGR	0.86 (1.43)	0.76 (1.07)	61.8	54.2	3.0	2.2	2.53	1.35	9.56	0.83	0.30	0.46
SV	0.30 (1.18)	0.14 (0.53)	88.1	90.9	1.1	0	5.36	4.26	33.02	18.87	0.34	0.23
ALK	0.38 (1.27)	2.54 (1.38)	84.7	18	0.9	18.5	4.85	-0.93	27.21	-0.55	0.28	0.04
KOG	1.16 (1.35)	0.67 (0.86)	47.0	48.7	4.9	0.7	1.15	1.41	2.25	1.76	0.28	0.23
KÖRP	1.36 (1.42)	0.88 (1.07)	42.9	47.4	6.8	0.2	0.72	0.95	0.47	-0.4	0.12	0.14
WAHN	0.17 (0.95)	0.07 (0.30)	92.6	94	0.3	0	7.62	5.56	63.96	36.37	0.27	0.19
DEP	2.83 (0.95)	1.70 (1.11)	3.1	15.3	21.9	3.6	-0.60	0.1	2.48	-0.77	0.29	0.33
AND	3.17 (0.94)	1.96 (1.32)	3.2	22.4	36.9	9.7	-0.93	-0.24	4.86	-1.16	0.27	0.47
BEZ	2.46 (1.31)	2.30 (1.14)	12.7	7.4	17.9	12.3	-0.12	-0.35	1.79	-0.74	0.28	0.56
AKT	1.81 (1.20)	1.54 (1.30)	20.6	27.1	4.5	7.7	0.05	0.34	0.53	-1.04	0.36	0.62
WOHN	0.76 (1.50)	1.51 (1.43)	66.6	34	1.9	9.8	3.11	0.41	12.70	-1.29	0.26	0.55
BERUF	1.98 (1.60)	2.15 (1.26)	26.5	15.2	11.0	7.6	0.85	-0.44	2.76	-1.0	0.27	0.53

Anmerkungen: r_{it} = Trennschärfe (korrigiert); SD = Standardabweichung; AGGR = Item 1 “Aggression”, SV = Item 2 “Selbstverletzung”, ALK = Item 3 “Alkohol- und Drogenmissbrauch”, KOG = Item 4 “Kognitive Probleme”, KÖRP = Item 5 “Körperliche Probleme”, WAHN = Item 6 “Halluzinationen und Wahn”, DEP = Item 7 “Depressivität”, AND = Item 8 “Andere Probleme”, BEZ = Item 9 “Probleme mit Beziehungen”, AKT = Item 10 “Probleme mit alltäglichen Aktivitäten”, WOHN = Item 11 “Probleme mit den Wohnbedingungen”, BERUF = Item 12 “Probleme in Beruf und Freizeit”

6.3. Reliabilität

Die Reliabilität der Items der HoNOS-D wurde ebenfalls bei zwei verschiedenen Stichproben untersucht. Zum einen wurde die Test-Retest-Reliabilität bei Patienten mit psychischen Erkrankungen in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung (Stichprobebeschreibung a, siehe Kap. 5) und die Interraterreliabilität an Patienten mit Abhängigkeitserkrankungen (Stichprobebeschreibung c, siehe Kap. 5) untersucht. Die Überprüfung der Interraterreliabilitäten des weiter entwickelten Items 8 der HoNOS-D wurde an der Stichprobe von Patientinnen mit psychischen Erkrankungen in stationärer Rehabilitationsbehandlung (siehe Stichprobebeschreibung b, siehe Kap. 5) vorgenommen.

6.3.1 Test-Retest-Reliabilität

Aus:

Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology*, 40, 116-125.

Erwartungsgemäß wiesen hinsichtlich der Test-Retest-Reliabilitäten 9 der 12 Items der HoNOS-D zufrieden stellende Werte auf (siehe Tabelle 6). Nur 3 der 12 Item wiesen nach den Kriterien von Landis und Koch (1977) unzufrieden stellende Werte auf: Item 3 „Alkohol- und Drogenmissbrauch“, Item 7 „Depressivität“ und Item 8 „Andere Probleme“. Das könnte damit zu tun haben, dass die Beurteiler in dieser Studie kein spezifisches Training im HoNOS-D erhalten haben. Im Fall von Item 8 „Andere Probleme“ könnte es auch an der fehlenden Spezifizierung im Glossar gelegen haben. Demzufolge haben wir in der weiter entwickelten Version des Items 8 der HoNOS-D Interraterreliabilitäten untersucht.

6.3.2 Interraterreliabilitäten

Aus:

Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.

Die Intraklassenkorrelationskoeffizienten für die verschiedenen Problembereiche, die unter Item 8 subsumiert sind, können der Tabelle 6 entnommen werden. Entsprechend den Kriterien

von Landis und Koch (1977) erreichen allerdings nur 3 der 10 Problembereiche zufrieden stellende Werte (κ zwischen 0,61 – 0,80): Die höchsten Interraterübereinstimmungen zeigen sich für den Bereich „Dissoziativ“ und „Essen“, überraschenderweise gefolgt von dem Bereich „Somatoform“ (siehe Tabelle 6). Dieses Ergebnis steht im Einklang mit empirischen Ergebnissen einer anderen Studie, in der die Validität der Diagnosestellung aus klinischen Routineinterviews untersucht wurde (Andreas et al., 2009c). Auch dort zeigte sich, dass diejenigen Diagnosen, die die auffälligsten Verhaltensweisen aufwiesen, auch zuverlässiger im klinischen Routineinterview erkannt wurden. So könnte es sich bei diesen Ergebnissen zur Interraterreliabilität der Problembereiche unter Item 8 um ein ähnliches Phänomen handeln. Patienten, die dissoziative oder essensbezogene Probleme aufweisen, sind anscheinend von unterschiedlichen Beurteilern, auch hinsichtlich ihres Schweregrades zuverlässiger einzuschätzen. Unzufrieden stellende Interraterreliabilitätskoeffizienten wiesen dagegen die drei Bereiche „Sexualität“, „Angst“ und „Stress“ auf. Hier wird ebenfalls sehr deutlich, dass ebenso wie für die anderen 11 Items der HoNOS-D ein Glossar mit Ankerbeispielen für die 10 Problembereiche des Items 8 der HoNOS-D erarbeitet und psychometrisch überprüft werden müsste. Damit wäre an dieser Stelle auf weiteren Forschungsbedarf hinzuweisen.

Aus:

Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.

In der Studie an **Patienten mit Abhängigkeitserkrankungen** wurden Interraterübereinstimmungen **vor** und **nach** einer spezifischen **Schulung** im HoNOS-D untersucht. Die Schulung erfolgte in Anlehnung an die von Wing et al. (1996) erarbeiteten Schulungsunterlagen und wurde um einige Bestandteile aus dem Schulungsprogramm Australiens (Hospitalisation bei Tom Trauer, Melbourne, Australien in 2004) ergänzt. Die Schulung dauerte insgesamt 3 Stunden und beinhaltete neben einer spezifischen Erklärung der Items der HoNOS-D mehrere Fallbeispiele, anhand derer das Rating praktisch geübt wurde.

Obgleich empirische Befunde von Rock und Preston (Rock & Preston, 2001) darauf hinweisen, dass für die Beurteilungen im HoNOS-D kein spezifisches Training notwendig ist, zeigen die Ergebnisse unserer Studie, dass 5 der 12 Items vor dem Training nicht zufrieden stellende Interraterreliabilitäten (IRR) aufweisen. Nach dem Training hingegen können für fast alle Items der HoNOS-D zufrieden stellende Werte aufgezeigt werden. Nur Item 9 „Probleme mit

Beziehungen“ wies auch nach dem Training keine zufrieden stellende IRR auf. Zu ähnlichen Ergebnissen gelangten auch die Studien von Brooks (2000), Orrell et al., (1999) und Trauer et al. (1999). Möglicherweise ist das Item 9 „Probleme mit Beziehungen“ eher schwieriger zu Beginn der Behandlung zu beurteilen, da im klinischen Erstinterview der Schwerpunkt mehr auf der Diagnosestellung und Therapieplanung liegt. Abschließend kann festgehalten werden, dass in einer weiteren Studie, die gerade in Vorbereitung ist, Trainings-, Rater- und Übungseffekte bei der Beurteilung im HoNOS-D untersucht werden sollen (Andreas et al., 2009b). Erste Ergebnisse weisen aber auch hier darauf hin, dass in der Mehrzahl der Items der HoNOS-D Trainingseffekte zu beobachten sind.

Tabelle 6: Intraklassenkorrelationskoeffizienten für die Test-Retest-Reliabilität und für die Interraterreliabilität der Items im HoNOS-D und für das weiter entwickelte Item 8

Items der HoNOS-D	aus (Andreas et al., 2007)	aus (Andreas et al., in press)		aus (Andreas et al., sub- mitted)
	<i>Test-Retest</i>	<i>Interraterreliabilität</i>		<i>Interraterreliabilität</i>
	ICC	ICC _{vor Schulung}	ICC _{nach Schulung}	ICC für Item 8 Wei- terentwicklung
AGGR	0.80	0.75	0.82	
SV	0.87	0.82	0.81	
ALK	0.73	0.76	0.78	
KOG	0.89	0.54	0.68	
KÖRP	0.91	0.51	0.89	
WAHN	0.91	0.11	0.71	
DEP	0.75	0.49	0.80	
AND	0.72	0.73	0.82	
<i>Phobisch</i>				0.47
<i>Angst</i>				0.37
<i>Zwang</i>				0.58
<i>Stress</i>				0.24
<i>Dissoziativ</i>				0.73
<i>Somatoform</i>				0.66
<i>Essen</i>				0.73
<i>Schlaf</i>				0.41
<i>Sexuell</i>				0.29
<i>Schweregrad</i>				0.34
<i>Auswahl des Prob- lems</i>				0.23
BEZ	0.84	0.65	0.64	
AKT	0.83	0.34	0.86	
WOHN	0.90	0.76	0.79	
BERUF	0.89	0.69	0.76	

Anmerkungen: ICC = Intraklassenkorrelationskoeffizienten; AGGR = Item 1 “Aggression”, SV = Item 2 “Selbstverletzung”, ALK = Item 3 “Alkohol- und Drogenmissbrauch”, KOG = Item 4 “Kognitive Probleme”, KÖRP = Item 5 “Körperliche Probleme”, WAHN = Item 6 “Halluzinationen und Wahn”, DEP = Item 7 “Depressivität”, AND = Item 8 “Andere Probleme”, BEZ = Item 9 “Probleme mit Beziehungen“, AKT = Item 10 “Probleme mit alltäglichen Aktivitäten“, WOHN = Item 11 “Probleme mit den Wohnbedingungen“, BERUF = Item 12 “Probleme in Beruf und Freizeit”

6.4. Validität

6.4.1 Faktorielle Validität

Aus:

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.

Auch die faktorielle Validität wurde an zwei unterschiedlichen Stichproben untersucht (siehe Stichprobenbeschreibung a) und c) in Kap. 5).

Zunächst wurde eine Interkorrelationsmatrix der Items der HoNOS-D berechnet. Diese zeigte erwartungsgemäß auf, dass von 66 möglichen Korrelationen zwischen den 12 Items nur 5 Korrelationen eine mindestens mittlere ES von $r > 0,30$ aufwiesen (Andreas et al., 2010).

Wie im Kap. 4 bereits ausführlich beschrieben, konnte die von Wing et al. (1998) postulierte Skalenstruktur, die inhaltlich festgelegt wurde, in anderen Studien nicht bestätigt werden. Im Folgenden soll deswegen skizziert werden, dass sowohl die von den Testautoren postulierte Originalstruktur als auch die Faktorenstruktur von Trauer (1999) und Eagar (2005) mittels Konfirmatorischer Faktorenanalysen überprüft wurden.

Zunächst wurde eine Überprüfung der a priori formulierten Skalenstruktur der Testautoren (Wing et al., 1998) vorgenommen. Dazu wurde eine Konfirmatorische Faktorenanalyse auf der Basis von polychorischen Korrelationsmatrizen berechnet. Die folgenden Werte konnten ermittelt werden: $\chi^2 = 455,69$ (df = 48, p = 0,001), RMSEA = 0,12, SRMR = 0,08 and CFI = 0,69. Da die gefundenen Werte außerhalb der von Hu und Bentler (Hu & Bentler, 1999) empfohlenen cut-off Scores lagen, konnte die Skalenstruktur der Testautoren nicht bestätigt werden. Auch die von Trauer (Trauer, 1999) entwickelte 5-Faktorenlösung konnte in unserer Studie nicht bestätigt werden.

Lauzon et al. (2001) untersuchte ein 1-dimensionales Modell der HoNOS mittels Konfirmatorischer Faktorenanalyse. In Anlehnung an dieses Vorgehen überprüften wir zwei 1-dimensionale Modelle: Das erste Modell umfasste die Items 1 bis 10 der HoNOS-D und das zweite Modell die Items 1 bis 12 der HoNOS-D. Das 1-dimensionale Modell mit 10 Items zeigte folgende Fitindizes auf: $\chi^2 = 394,11$ (df = 35, p = 0,001), RMSEA = 0,13, SRMR =

0,09 and CFI = 0,63, die sich ebenfalls alle außerhalb der empfohlenen Grenzen von Hu und Bentler (Hu & Bentler, 1999) befanden. Auch das 1-dimensionale Modell mit 12 Items zeigte ähnliche Ergebnisse.

Eine entsprechende Prüfung der internen Konsistenz des Gesamtscores (über 10 Items und über 12 Items) wies unzufrieden stellende Werte auf (Cronbach's α 10 Items = 0,60 und für 12 Items = 0,62).

Wir überprüften ebenfalls die von den Testautoren postulierte Skalenstruktur an der Stichprobe bei **Patienten mit Abhängigkeitserkrankungen** (Andreas et al., in press). Auch hier zeigte die Interkorrelationsmatrix, dass nur 15 von 66 möglichen Interkorrelationen ES über $r > 0,30$ aufwies. Dagegen konnte die mittels KFA ermittelten Fitindizes zufrieden stellende Werte ermitteln: RMSEA = 0,09; SRMR = 0,07; CFI = 0,97; TLI = 0,95). Wir konnten die Skalenstruktur der Testautoren an dieser Stichprobe bestätigen. Dennoch blieben die internen Konsistenzen für 3 der 4 Subskalen in einem unzufrieden stellenden Bereich (Skala 1 "Verhalten" $\alpha = 0,05$, Skala 2 "Impairment" $\alpha = 0,20$, Skala 3 "Symptome" $\alpha = 0,49$ und Skala 4 "Soziale Probleme" $\alpha = 0,82$). In dieser Studie wies dagegen der Gesamtscore (über alle 12 Items) mit $\alpha = 0,73$ eine zufrieden stellende Höhe auf.

Basierend auf diesen Ergebnissen nahmen wir im Folgenden eine Validitätsprüfung auf Einzelitemebene vor. Aus inhaltlichen Gründen schlossen wir in die Validitätsprüfung auch Untersuchungen zum Gesamtscore über 12 Items der HoNOS-D ein.

6.4.2 Konvergente und diskriminante Validität

- a) bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

Bezüglich der konvergenten Validität der Items der HoNOS-D zeigte sich, dass erwartungsgemäß Items mit den entsprechenden Subskalen der anderen eingesetzten Instrumente korrelierten (z.B. Item 7 „Depressivität“ mit der Depressivitätsskala der SCL-14 oder Item 5 „Körperliche Probleme“ mit der Körperlichen Summenskala des SF-8-F). Die Höhe der Korrelati-

onen entsprach ebenfalls empirischen Befunden anderer Studien (Adams et al., 2000). Hier sollte aber darauf hingewiesen werden, dass die anderen Studien eine Validierung des Gesamtscores der HoNOS-D und nicht der Einzelitems vorgenommen haben.

Überraschenderweise konnten aber für einige erwartete Zusammenhänge keine Bestätigung gefunden werden. So wurde z.B. kein Zusammenhang mit mindestens mittlerer ES zwischen dem HoNOS-D Item 9 „Probleme mit Beziehungen“ und dem Gesamtscore des IIP gefunden werden. Ein möglicher Grund könnte darin liegen, dass das Item 9 „Probleme mit Beziehungen“ eher allgemein auf die Aufrechterhaltung von Beziehungen bezogen ist, während das IIP spezifische interaktionelle Verhaltensweisen abfragt. Ebenfalls entgegen den empirischen Ergebnissen anderer Studien konnte in dieser Studie kein Zusammenhang mit mindestens großer ES zwischen den HoNOS Items und dem GAF Score gefunden werden (vgl. hierzu Kap. 4). Das könnte auch damit verbunden gewesen sein, dass in den Videos, die als Grundlage zur Bewertung der HoNOS dienten, kaum Informationen zur sozialen Funktionsfähigkeit enthalten waren und deswegen die korrespondierenden Items der HoNOS-D (z.B. Item 10 „Probleme mit alltäglichen Aktivitäten“ oder Item 11 „Probleme mit den Wohnbedingungen“) nur schwer einschätzbar waren.

b) Item 8 Weiterentwicklung bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.

Für das weiter entwickelte Item 8 der HoNOS-D wurden ebenfalls Validitätsprüfungen vorgenommen. Hier zeigten sich signifikante Korrelationen mit annähernd mittlerer ES zwischen den Problembereichen „Phobisch“ und „Angst“ des Items 8 und der Subskala „Phobische Angst“ der SCL-14. Mit einer mindestens kleinen ES korrelierte der Problembereich „Somaform“ mit der Somatisierungsskala der SCL-14 ($r = 0,23$). Keine Korrelationen mit mindestens kleiner ES zeigten sich für die Problembereiche „Zwanghaft“ und „Sexuell“ und dem GSI der SCL-14. Einige signifikante Zusammenhänge mit mindestens mittlerer ES konnten auch zwischen den Problembereichen des weiter entwickelten Items 8 und den beiden Subskalen „Körperliche und Psychische Summenskala“ des SF-8-F nachgewiesen werden. Über-

raschenderweise zeigte sich hierbei eine signifikante Korrelation mit $r = 0,26$ zwischen dem Problembereich „Angst“ und der „Psychischen Summenskala“ des SF-8-F. Dabei kann durchaus vermutet werden, dass eine Beeinträchtigung durch Probleme, die im Zusammenhang mit Angst stehen, auch mit einer Beeinträchtigung der psychischen Lebensqualität assoziiert sind.

Kritisch muss an dieser Stelle angemerkt werden, dass obwohl einige Zusammenhänge erwartungsgemäß aufgezeigt werden konnten, aus pragmatischen Gründen keine spezifischen Instrumente zur Validierung der verschiedenen Problembereiche des Items 8 (wie z.B. „Sexuell“ oder „Essen“) zum Einsatz kamen. An dieser Stelle kann sicherlich noch weiterer Forschungsbedarf verzeichnet werden.

6.4.3 Kriteriumsvalidität

- a) bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

Basierend auf der Stichprobe b) (siehe Kap. 5) wurde im Sinne der Kriteriumsvalidität überprüft, ob Indexitems der HoNOS-D (z.B. Item 7 „Depressivität“) in ihrer Ausprägung zwischen Patienten mit der korrespondierenden Diagnose (z.B. Depression – F32-34 nach ICD-10) und ohne diese Diagnose differenzieren können. Dabei konnte aufgezeigt werden, dass die kriteriumsbezogene Validität für alle Indexitems bestätigt wurde. So konnte zum Beispiel auch gezeigt werden, dass das Indexitem 3 „Alkohol- und Drogenmissbrauch“ sich in seinem Schweregrad signifikant zwischen Patienten mit und ohne substanzbezogener Diagnose unterschied. Es zeigte sich aber auch, dass andere Items der HoNOS-D (z.B. Item 8 „Andere Probleme“ oder Item 2 „Selbstverletzendes Verhalten“) zwischen Patienten mit und ohne substanzbezogene Störung differenzierten. Das könnte damit zusammenhängen, dass Patienten mit Abhängigkeitserkrankungen auch eine höhere Komorbidität aufweisen und demzufolge in mehreren Problembereichen Beeinträchtigungen zeigen (Ilgen et al., 2005).

b) Item 8 Weiterentwicklung bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.

Auch für das weiter entwickelte Item 8 wurde eine Überprüfung der Kriteriumsvalidität an der Stichprobe b) (siehe Stichprobenbeschreibung Kap. 5) vorgenommen. Auch hier konnten für die meisten Problembereiche (= Indexitems) Unterschiede in der Ausprägung der korrespondierenden Diagnosen gefunden werden. So zeigte sich für Patienten mit Angststörungen, dass die Problembereiche „Phobisch“ und „Angst“ zwischen dem Vorhandensein bzw. dem Nichtvorhandensein einer Angststörung diskriminierten. Das zeigte sich ebenso für den Problembereich „Somatoform“ für Patienten mit und ohne somatoforme Störung, für den Problembereich „Dissoziativ“ für Patienten mit und ohne PTSD (Posttraumatische Belastungsstörung) und für den Problembereich „Appetit/Essen“ für Patienten mit/ohne Essstörung. In dieser Studie fanden wir auch heraus, dass nicht immer die Indexitems die höchsten Ausprägungen und demzufolge Schweregrade bei der korrespondierenden Diagnose aufwiesen. So fand sich relativ konsistent, dass der Problembereich „Stress“ in allen untersuchten Diagnosen eine hohe Ausprägung im Schweregrad aufwies. Das könnte damit zu tun haben, dass es sich hierbei um einen generellen Belastungsfaktor handelt, der unabhängig von der Diagnose auftritt und eventuell auch das Leiden durch die Erkrankung repräsentiert. Damit konnten wir auch aufzeigen, dass das differenzierte Rating des weiter entwickelten Items 8 auch zu einem differenzierten Bild der Problembereiche bei den Patienten führte, womit gerade für diesen Versorgungsbereich bei Einsatz der Originalversion wesentliche Information verloren gehen würden.

6.4.5 Änderungssensitivität

a) bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, 19, 50-62.

Bezüglich der Änderungssensitivität fanden wir mindestens kleine Effektstärken im Unterschied zwischen Beginn und Ende der Behandlung in 9 der 12 Items der HoNOS-D und mindestens mittlere Effekte ($d = 0,50 - 1,34$) zeigten sich für 4 der 12 Items (Item 7 „Depressivität“, Item 8 „Andere Probleme“, Item 9 „Probleme in Beziehungen“ und Item 10 „Probleme mit alltäglichen Aktivitäten“, siehe Tabelle 7). Keine Unterschiede zwischen Beginn und Ende der Behandlung fanden sich im Item 2 „Selbstverletzung“, Item 6 „Halluzinationen und Wahn“ und Item 11 „Probleme mit den Wohnbedingungen“. Das lässt sich dadurch erklären, dass Item 2 und Item 6 bereits zu Beginn der Behandlung kaum eine Relevanz für Patienten in stationärer psychosomatischer/psychotherapeutischer Rehabilitationsbehandlung aufweisen. Item 11 bezieht sich auf die Einschätzung der Wohn- und Lebensbedingungen des Patienten und lässt somit auch keine Veränderungen während einer stationären Behandlung, die im Durchschnitt 6 Wochen dauert, erwarten.

c) Item 8 Weiterentwicklung bei Patienten mit psychischen Erkrankungen in psychosomatischer/psychotherapeutischer Behandlung

Aus:

Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.

In der Beurteilerinstruktion für Item 8 der Originalversion werden die Beurteiler gebeten, dass schwerste klinische Probleme der letzten 2 Wochen einzuschätzen. Am Ende der Behandlung könnte sich allerdings dieses Problem eben durch die Behandlung selbst, geändert haben (z.B. zu Beginn der Behandlung „Phobisch“ und am Ende der Behandlung „Schlaf“). Wenn es einen solchen Wechsel geben sollte, so gäbe es zum Ende der Behandlung in der Originalversion des Items 8 keine Möglichkeit die Veränderung zum Ausgangszustand festzuhalten. Deswegen kann an dieser Stelle schon konstatiert werden, dass die Änderungssensitivität für Item 8 der Originalversion möglicherweise eingeschränkt ist.

Ohne Berücksichtigung des möglichen Wechsels des Problems konnte für Item 8 der Originalversion eine ES von $d = 0,98$ aufgezeigt werden. Für die weiter entwickelte Version des Items 8 – wieder ohne Berücksichtigung des Wechsels der Problematik zum Ende der Behandlung – zeigte sich eine ES von $d = 1,24$. Bei Vergleich der Veränderung im Schweregrad

bezogen auf das gleiche Problem zu Beginn und am Ende der Behandlung in der weiter entwickelten Version des Items 8 konnte eine ES von $d = 1,44$ nachgewiesen werden.

Tabelle 7: Änderungssensitivität der HoNOS-D Items im Vergleich zu anderen Outcomeinstrumenten

	zu Beginn	am Ende	p^1		ES
	$M_1 (SD_1)$	$M_2 (SD_2)$			
HoNOS-D					
AGGR	0.76 (1.16)	0.37 (0.76)	0.001		0.35
SV	0.20 (0.72)	0.10 (0.44)	0.001		0.15
ALK	0.25 (0.76)	0.09 (0.44)	0.001		0.21
KOG	1.13 (1,27)	0.69 (1.00)	0.001		0.35
KÖRP	1.33 (1.37)	0.90 (1.14)	0.001		0.32
WAHN	0.09 (0.47)	0.05 (0.35)	0.001		0.08
DEP	2.83 (0.93)	1.57 (1.00)	0.001		1.34
AND	3.17 (0.90)	2.14 (0.96)	0.001		1.14
BEZ	2.44 (1.23)	1.81 (1.16)	0.001		0.50
AKT	1.80 (1.18)	0.94 (1.02)	0.001		0.73
WOHN	0.62 (1.07)	0.46 (0.91)	0.001		0.15
BERUF	1.89 (1.39)	1.34 (1.24)	0.001		0.40
Tot12	16.99 (6.31)	10.86 (6.02)	0.001		0.97
SCL					
DEP	1.77 (1.04)	1.10 (0.95)	43.24	3168	0.65
PHOB	0.53 (0.79)	0.31 (0.61)	22.50	3168	0.29
SOM	1.23 (0.99)	0.84 (0.89)	27.72	3168	0.39
GSI	1.26 (0.74)	0.80 (0.69)	44.03	3168	0.63
SF-8					
PSK-S	2.23 (0.88)	1.47 (0.93)	47.73	3168	0.86
PSK-F	2.60 (0.66)	1.36 (0.73)	88.26	3044	1.88
KSK-S	1.86 (0.92)	1.30 (0.88)	39.69	3168	0.60
KSK-F	1.91 (0.81)	1.06 (0.68)	66.66	3037	1.04

Anmerkungen: ¹ Wilcoxon rank sum tests; all t-values are significant ($p < 0.001$). AGGR = Item 1 "Aggression", SV = Item 2 "Selbstverletzung", ALK = Item 3 "Alkohol- und Drogenmissbrauch", KOG = Item 4 "Kognitive Probleme" KÖRP = Item 5 "Körperliche Probleme", WAHN = Item 6 "Halluzinationen und Wahn", DEP = Item 7 "Depressivität", AND = Item 8 "Andere Probleme", BEZ = Item 9 "Probleme mit Beziehungen", AKT = Item 10 "Probleme mit alltäglichen Aktivitäten", WOHN = Item 11 "Probleme mit den Wohnbedingungen", BERUF = Item 12 "Probleme in Beruf und Freizeit", Tot12 = total score (12 items) of HoNOS, SCL = Symptom-Check-List-90-R, SCL-DEP = Depressionsskala des SCL, SCL-PHOB = Phobische Subskala, SCL-SOM = Somatoforme Subskala, SCL-GSI = Gesamtbelastungsindex der SCL, SF-8 = Short-Form-8, SF8-PSK-S = Short-Form-8-Psychische-Summenskala-Selbstrating, SF8-KSK-S = Short-Form-8-Körperliche Summenskala Selbstrating, SF8-PSK-F = Short-Form-8- Psychische-Summenskala Fremdrating, SF8-KSK-F = Short-Form-8- Körperliche Summenskala - Fremdrating ES = Effektstärke nach Cohen's d ($d < 0.20$ = kleiner Effekt, $d = 0.20 - 0.50$ mittlerer Effekt and $d > 0.80$ großer Effekt)

6.4.6 Prognostische Validität

Aus:

Andreas, S., Dirmaier, J., Harfst, T., Kawski, S., Koch, U., & Schulz, H. (2009). On the development and evaluation of a case group concept for inpatients with mental disorders: Using self-report and expert-rated instruments. *European Psychiatry*, 24, 105-111.

Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.

Die prognostische Validität der HoNOS-D wurde an zwei Stichproben untersucht (siehe Stichprobenbeschreibung a) und c) im Kap. 5). Dazu wurde in beiden Studien ein Regression-Tree-Verfahren verwendet, um die Vorhersagekraft des Prädiktors „Schweregradeinschätzung im HoNOS-D“ zum einen bezüglich der Behandlungsdauer (Andreas et al., 2009a) und zum anderen bezüglich des Ressourcenverbrauches in stationärer Rehabilitationsbehandlung bei Patienten mit Abhängigkeitserkrankungen zu untersuchen (Andreas et al., in press). Da in der Studie von Andreas et al. (in press) explizit der Fokus auf der Untersuchung der prognostischen Validität lag und nicht auf der Entwicklung eines Klassifikationssystems, sollen im Folgenden die Ergebnisse dieser Studie etwas ausführlicher dargestellt werden.

In beiden Studien erwies sich entweder der Gesamtscore der HoNOS-D bzw. die Einzelitems (Item 1 „Aggressivität“) als relevante Prädiktoren des Ressourcenverbrauches in stationärer Behandlung (siehe Anhang 5 bzw. 6).

In der Studie von Andreas et al. (in press) wurden weitere Variablen in das Modell eingespeist: Geschlecht, Alter zu Beginn, Schulabschluss, Berufsstatus, Familienstand, substanzbezogene Störung als Hauptdiagnose, komorbide Störung und somatische Komorbidität. Von allen eingespeisten Prädiktoren sagten 10 Merkmale signifikant den Gesamtressourcenverbrauch der KTL-Leistungen in stationärer Rehabilitationsbehandlung vorher. Diese erklärten insgesamt 58% der Gesamtvarianz am Ressourcenverbrauch. Item 1 „Aggressivität“ wurde in dieser Analyse als der bedeutsamste Prädiktor identifiziert. Auf der zweiten Ebene des Entscheidungsbaumes erwiesen sich folgende Merkmale als relevant: „Schulabschluss“, „Persönlichkeitsstörung (F60 nach ICD-10) und „Berufsstatus“. Weiterhin sagten die Variablen „Familienstand“, „Geschlecht“, „multiple substanzbezogene Störung (F19 nach ICD-10)“, „Alter bei Aufnahme“, „cannaboidbezogene Störung (F12 nach ICD-10)“ und Item 3 „Alkohol- und Drogenmissbrauch“ den Ressourcenverbrauch in stationärer Rehabilitationsbehandlung signi-

signifikant vorher. Konsistent mit früheren Arbeiten und empirischen Ergebnissen anderer Studien zeigte sich auch in dieser Studie, dass Patienten mit einem höher ausgeprägten Schweregrad auch einen höheren Ressourcenverbrauch aufwiesen (Ashcraft et al., 1989; Buckingham et al., 1998; Gaines et al., 2003; Horn et al., 1989; Mitchell et al., 1987; Taube et al., 1984b). Dabei hatten diejenigen Patienten mit dem höchsten Ressourcenverbrauch auch eine höhere Ausprägung im Item 1 „Aggressivität“, einen höheren Schulabschluss und waren nicht verheiratet ($M = 6159$ min, $SD = 2239$ min, $n = 69$). Im Gegensatz dazu wiesen diejenigen Patienten mit dem niedrigsten Ressourcenverbrauch auch eine niedrigere Schweregradausprägung im Item 1 „Aggressivität“ der HoNOS-D auf und befanden sich in beruflicher Rehabilitation ($M = 61$ min, $SD = 343$ min, $n = 32$, siehe Entscheidungsbaum in der Studie Andreas et al., in press im Anhang 5). Ein zweites Entscheidungsbaummodell mit dem Gesamtscore der HoNOS-D als Prädiktor ergab ähnliche Resultate.

Obgleich die extrahierte Gesamtvarianz mit 58% im Vergleich zu Ergebnissen eigener Arbeiten (Andreas et al., 2009a) und anderer Studien (Ashcraft et al., 1989; Buckingham et al., 1998; Gaines et al., 2003; Horn et al., 1989; Mitchell et al., 1987; Taube et al., 1984b) als relativ substantiell zu bewerten ist, was ein Hinweis darauf ist, dass die ausgewählten Prädiktoren auch relevante Bereiche des Ressourcenverbrauches vorhersagen können, bedarf es dennoch weiterer Studien, die andere, den Ressourcenverbrauch beeinflussende Faktoren erfassen (z.B. Krankenversicherung, Privat vs. Gesetzlich). Weiterhin sollten nachführende Studien untersuchen, ob Patienten mit Abhängigkeitserkrankungen auch auf der Systemebene die für ihre Behandlung notwendigen Ressourcen erhalten (z.B. im ambulanten Sektor).

7. Zusammenfassung

Im folgenden Kapitel sollen die gerade berichteten Ergebnisse zusammenfassend hinsichtlich der in Kap. 4 aufgeworfenen Fragen bewertet werden. Im nachfolgenden Kapitel sollen dann Implikationen für Forschung und Praxis beschrieben werden.

Die hier zusammengetragenen Studien wurden dabei von folgenden Fragestellungen, die nun zusammenfassend beantwortet werden sollen, geleitet:

- 1. Wie viel Aufwand ist nötig, damit der HoNOS eingeschätzt werden kann? Wie viele gezielte Fragen sind über ein allgemeines klinisches Interview hinausgehend dazu nötig?**

Wir konnten zeigen, dass erfahrene Beurteiler zum Ausfüllen der HoNOS-D im Durchschnitt 5 Minuten benötigen. Auch die Analyse fehlender Werte zeigte für die Mehrzahl der Items zufrieden stellende Ergebnisse. Damit kann konstatiert werden, dass der HoNOS-D sich in den hier untersuchten Versorgungsbereichen (Patienten mit psychischen Erkrankungen in stationärer psychosomatischer/psychotherapeutischer Behandlung und Patienten mit Abhängigkeitserkrankungen in stationärer Rehabilitationsbehandlung) als weitest gehend praktikabel erwiesen hat. Ein in der Praktikabilitätsanalyse dabei als problematisch, aber als sehr relevant identifiziertes Item, wurde von uns in der Form weiter entwickelt, dass es signifikant weniger fehlende Werte aufwies als die Originalversion (Item 8 „Andere psychische und verhaltensbezogene Probleme“).

Die Informationen zur Beurteilung im HoNOS-D können größtenteils aus dem klinischen Erstgespräch entnommen werden, welches die aktuelle Befindlichkeit (bzw. die Befindlichkeit der letzten Woche vor Aufnahme), Verhaltensbeobachtungen, Beziehungen und die aktuelle Situation im Beruf oder im Wohnen des Patienten erfasst. Sollten darüber hinaus weitere Informationen benötigt werden, so ist es den Testautoren zufolge auch möglich, alle an der Behandlung des Patienten beteiligten Berufsgruppen (z.B. Sozialarbeiter oder Krankengymnasten für eine Einschätzung im Item 5 „Körperliche Probleme) um Informationen zu bitten (Wing et al., 1996). Wir konnten diesen Aspekt ebenso in einer eigenen Studie nachweisen (Andreas et al., 2010). Anhand von standardisierten, strukturierten diagnostischen Interviews von Patienten zu Beginn einer psychosomatischen/psychotherapeutischen Rehabilitationsbe-

handlung (siehe Stichprobenbeschreibung b), Kap. 5), die auf Video aufgezeichnet wurden und drei im HoNOS-D geschulten Beurteilern vorgelegt wurden, konnte auf der Basis der diagnostischen Information ein Rating im HoNOS-D durchgeführt werden. Dabei zeigte sich, dass Item 10 „Probleme mit alltäglichen Aktivitäten“, Item 11 „Probleme mit den Wohnbedingungen“ und 12 „Probleme in Beruf und Freizeit“ die meisten fehlende Werte aufwies (Item 10 mit 14%, Item 11 mit 14% und Item 12 mit 32%). Im Rahmen eines klinischen Aufnahmegespräches wäre es aber vermutlich möglich gewesen, die Patienten nach den fehlenden Informationen zu fragen. Das war aufgrund der Videoaufzeichnungen in dieser Studie leider nicht möglich.

Es kann aber somit geschlussfolgert werden, dass sich der HoNOS-D in den hier untersuchten Versorgungsbereichen als ein weitest gehend praktikables und in der Routineversorgung anwendbares Instrument erwiesen hat.

2. Braucht es eine Schulung im HoNOS, damit er von den behandelnden Therapeuten ausgefüllt werden kann?

Die Testautoren wollten ein Instrumentarium konstruieren, welches in verschiedenen Settings der Versorgung von Patienten mit psychischen Erkrankungen und in der Routineversorgung verwendbar sein sollte. Damit ist natürlich die Frage verbunden, ob es einer Schulung bedarf, wenn das Instrument zum Beispiel im Rahmen von Qualitätssicherungsmaßnahmen großflächig eingesetzt werden soll. Empirische Befunde von Rock und Preston (2001) zeigten auf, dass es keinen signifikanten Unterschied in der Beurteilung der HoNOS in Abhängigkeit einer Schulung gab. Unsere ersten Auswertungen weisen darauf hin, dass eine Schulung im HoNOS notwendig ist. Wir führten dazu eine quasi-experimentelle Studie in vier verschiedenen Einrichtungen des Versorgungsbereiches Sucht durch (siehe Stichprobenbeschreibung c, Kap. 5). Dazu schätzten jeweils 2 Beurteiler pro Einrichtung einen Patienten bei Aufnahme ein. Insgesamt beurteilten 12 Therapeuten 281 Patienten mit Abhängigkeitserkrankungen. Nach 60 Ratings erfolgte eine ausführliche Schulung im HoNOS-D. Anschließend wurden ebenfalls noch einmal 60 Ratings in Raterpaarungen durchgeführt. Uns interessierte dabei insbesondere, ob ein Trainingseffekt vor bzw. nach Schulung nachgewiesen werden konnte oder ob eventuell Rater- bzw. Übungseffekte einen Einfluss auf die Beurteilung im HoNOS-D aufwiesen. Die Auswertung erfolgte mittels Hierarchisch-Linearem Modell (HLM) (Bryk & Raudenbush, 1992). Dabei fanden wir heraus, dass sich über alle Items Schulungseffekte zeig-

ten, wenn das Ausgangsniveau der Beurteiler und der Einrichtungen berücksichtigt wurde. Unsere ersten Analysen legen demzufolge eine Schulung im HoNOS-D nahe. In Australien, wo der HoNOS-D ein Instrument innerhalb eines Instrumentensets zur Qualitätssicherung von Behandlung bei Patienten mit psychischen Erkrankungen ist, wird seit Jahren flächendeckend in Form von Train-the-Trainer Seminaren Schulungen im HoNOS durchgeführt.

Sollte jedoch aus pragmatischen und ökonomischen Gründen eine Schulung im HoNOS-D nicht möglich sein, so haben wir eine Beurteilerinstruktion entwickelt, die auf einer Seite übersichtlich die wichtigsten Punkte aus dem Schulungsprogramm enthält. Dazu steht jedoch noch weiterer Forschungsbedarf aus. Hier könnte noch experimentell untersucht werden, welchen Nutzen die Beurteilerinstruktion hat und ob sie eine Überlegenheit gegenüber der Schulung aufweist.

3. Handelt es sich bei dem HoNOS um eine Sammlung von Einzelskalen oder können Subskalen bzw. ein Gesamtscore verwendet werden?

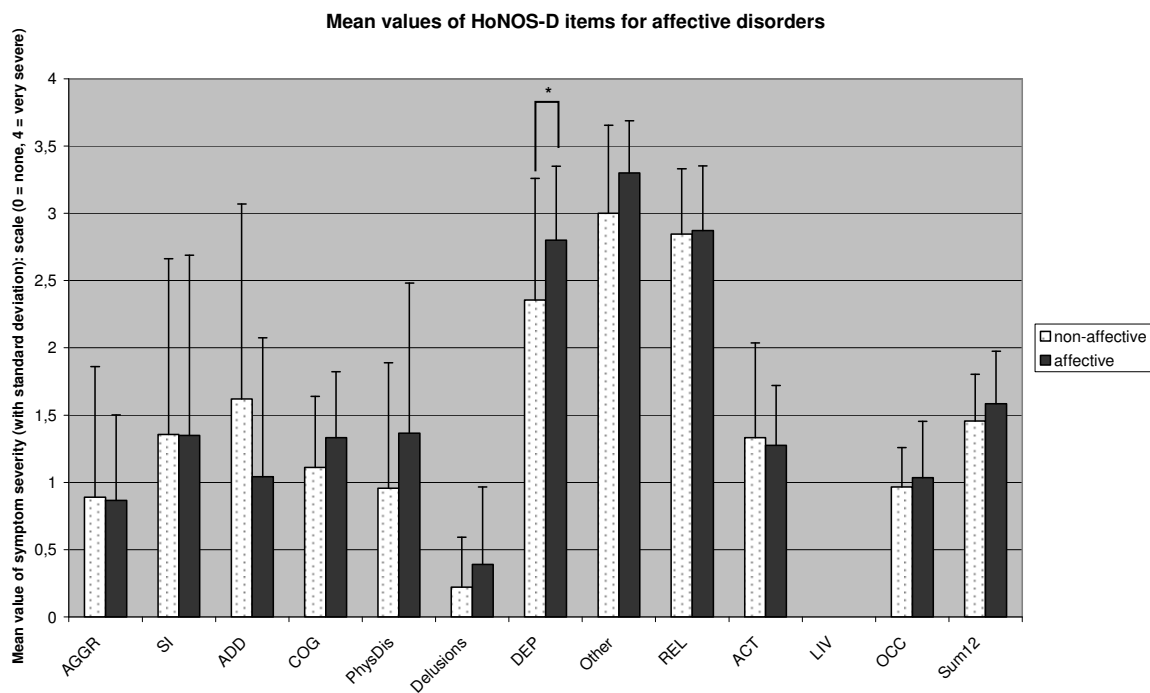
Entgegen der empirischen Befunde anderer Studien, die zwar teilweise niedrige interne Konsistenzen der Subskalen und des Gesamtscores der HoNOS berichten, aber dennoch Validierungen mit Subskalen bzw. mit dem Gesamtscore durchgeführt haben, fanden wir keine Bestätigung der Subskalen und des Gesamtscores des HoNOS-D. Auch explorative Faktorenanalysen konnten an unseren Stichproben keine zuverlässige und stabile Faktorenstruktur aufzeigen. Deswegen haben wir eine Reliabilitäts- und Validitätsüberprüfung anhand der Einzelitems der HoNOS-D vorgenommen. In Ergänzung und aus dem inhaltlichen Grund, dass der Gesamtscore als Gesamtschwerescore interpretierbar ist, haben wir ebenfalls unsere empirischen Analysen mit dem Gesamtsummenscore vorgenommen. Zusammenfassend kann geschlossen werden, dass sich sowohl die Einzelitems als auch der Gesamtscore der HoNOS-D als weitest gehend valide erwiesen.

Damit würden wir eine Verwendung der Einzelitems bzw. des 12-Items Gesamtscores in der Routineversorgung von Patienten mit psychischen Erkrankungen empfehlen. Weiterhin schätzen wir eine Verwendung der Subskalen als nicht aussagekräftig ein. Damit kann der HoNOS-D relativ schnell einen Überblick im Sinne einer Profildarstellung für die verschiedenen Problembereiche eines Patienten und dem klinisch praktisch tätigen Therapeuten ein differen-

ziertes Bild über verschiedene Problembereiche des Patienten (Aggressivität, Suizidalität oder Depressivität) geben.

4. Kann der HoNOS bei unterschiedlichen Patientengruppen und in unterschiedlichen Settings der Versorgung den Schweregrad differenziert messen?

Wie unsere verschiedenen Analysen aufzeigen, können die Items der HoNOS-D und auch der Gesamtscore als Schweregradausprägung in unterschiedlichen Problembereichen zwischen den Patienten, die sich in stationärer Behandlung befinden, zu Beginn differenzieren. Das zeigen zum Beispiel die empirischen Befunde zur Schweregradausprägung des Indexitems bei korrespondierenden Diagnosen (siehe Abbildung 3). Dieses Bild zeigt sich auch für das Indexitem 3 „Alkohol- und Drogenmissbrauch“ bei Patienten mit bzw. ohne substanzbezogene Störung (vgl. Anhang 4).



Mann Whitney U Test * $p < .05$ (Bonferroni-Holm adjusted)

Abbildung 3: Durchschnittliche Ausprägungen in den Einzelitems und im Gesamtscore der HoNOS-D bei Patienten mit bzw. ohne affektive Störung (aus Andreas et al., 2010).

Auch die empirischen Befunde zur Vorhersagekraft der Einzelitems und des Gesamtscores der HoNOS-D für die Behandlungsdauer bzw. den Ressourcenverbrauch von Patienten mit psychischen Erkrankungen in unterschiedlichen Settings zeigen auf, dass der HoNOS-D hinsichtlich der Schweregradausprägung zwischen verschiedenen Patientengruppen unterscheiden kann. So fanden wir, dass diejenigen Patienten mit den höchsten Schweregradausprägungen zu Beginn der Behandlung auch die längsten Behandlungsdauern bzw. den höchsten Ressourcenverbrauch aufwiesen (Andreas et al., 2009, Andreas et al., in press).

Somit kann geschlussfolgert werden, dass der HoNOS-D sich zufrieden stellend dazu eignet, zwischen Patienten mit unterschiedlichen Schweregradausprägungen zu unterscheiden.

5. Wie sehr schützt eine indirekte Ergebnismessung über wiederholte Beurteilungen im HoNOS-D vor dem Bias der direkten Bewertung der eigenen Arbeit?

Das ist eine schwierig zu beantwortende Frage. Autoren von anderen Studien (Hirdes et al., 2003), die sich mit der Vorhersage von Patientenmerkmalen in Hinblick auf den Ressourcenverbrauch beschäftigt haben, führen in diesem Zusammenhang auf, dass möglicherweise Anreize geschaffen werden, für die Behandlung von schwerer gestörten Patienten auch eine höhere finanzielle Vergütung zu erhalten. Das würde für eine Fremdbeurteilung des Schweregrades bedeuten, dass Therapeuten dazu neigen könnten, ihre eigene Arbeit besser einzuschätzen und damit am Ende der Behandlung niedrigere Ausprägungsgrade in einem Fremdbeurteilungsinstrument vergeben würden, als das tatsächlich der Fall wäre. Solche Effekte können bei Fremdbeurteilungsinstrumenten im Allgemeinen nur schwer kontrolliert werden. Eine Möglichkeit diesen Effekt zu verhindern, könnte darin bestehen, den Therapeuten gegenüber seiner Eingangsbeurteilung „blind“ zu machen. Im Falle unserer Studie hatte der Therapeut zu Ende der Behandlung nicht unbedingt Einblick in seine Einschätzung zu Beginn der Behandlung. Anhand des „Problemwechsels“ im Item 8 kann man zeigen, dass wohl ein Großteil der Therapeuten ihre Beurteilungen im HoNOS-D zu Beginn der Behandlung nicht verfügbar hatten (sonst hätten sie der Anweisung gemäß, keinen Wechsel im Problembereich des Items 8 vornehmen müssen).

Nichtsdestotrotz ist diese Frage auf der Basis der vorliegenden Datenlage nicht befriedigend zu beantworten. Es bedürfte dazu einer weiteren Studie, die einen „unabhängigen“ Beurteiler zum Ende der Behandlung die Einschätzung im HoNOS-D vornehmen lässt.

6. Kann das Item 8 „Andere Probleme“ die Komorbidität von psychischen Problemen praktikabel, reliabel und valide erfassen?

In der Originalversion ist das Item 8, welches andere psychische und verhaltensbezogene Probleme erfassen soll, vor dem Hintergrund der Analyse fehlender Werte als ein nicht ausreichend praktikables Item einzuschätzen. Da dieses Item allerdings sehr relevante Problembereiche bei Patienten mit psychischen Erkrankungen (wie z.B. Angst oder Zwang) erfasst, entwickelten wir dieses Item weiter. Die Weiterentwicklung erfolgte jedoch unter der Prämisse, dass eine Vergleichbarkeit zur Originalversion erhalten blieb (siehe dazu Anhang 1). Unsere Ergebnisse zeigen, dass sich das weiter entwickelte Item 8 gegenüber der Originalversion hinsichtlich der Praktikabilität überlegen zeigt. Eine leichte Überlegenheit gegenüber der Originalversion zeigte sich auch in der Validität der nun differenzierteren Einschätzung im Item 8. Da die Ergebnisse zur Reliabilität noch mehrheitlich unbefriedigend blieben, muss an dieser Stelle weiterer Entwicklungsbedarf konstatiert werden.

Es kann demzufolge geschlussfolgert werden, dass die Weiterentwicklung des Items 8 der HoNOS-D als ein wichtiger Schritt zur Verbesserung des Instrumentariums im Sinne auch der Erfassung von komorbiden Problembereichen angesehen werden kann. Es sollte aber auch unbedingt betont werden, dass es eines differenzierten Glossars wie für die anderen Items der HoNOS geben müsste.

8. Implikationen für Versorgung und Forschung

8.1. Klinische Implikationen

Wir konnten aufzeigen, dass der HoNOS-D ein für den Versorgungsbereich von Patienten mit psychischen Erkrankungen weitest gehend relevantes Fremdeinschätzungsinstrumentarium zur differenzierten Beurteilung des Schweregrades darstellt. Die hier aufgezeigten Ergebnisse können damit auch Hinweise für eine mögliche, zukünftigere Verbesserung sowohl in der Diagnostik des Schweregrades psychischer Erkrankungen als auch in der Versorgung von Patienten mit psychischen Störungen geben.

Darauf aufbauend wären folgende *Möglichkeiten* zur Verbesserung der klinischen Versorgung denkbar:

- Eine – wie auf der Basis der Ergebnisse vorgenommene differenzierte Beurteilung des Schweregrades von verschiedenen Problembereichen - könnte zur Optimierung indikativer Therapiezuweisung beitragen. So wäre denkbar, dass Patienten nicht nur aufgrund ihrer Diagnose zu gruppentherapeutischen Angeboten zugewiesen werden, sondern auch aufgrund von verschiedenen Schweregraden zu einheitlichen Gruppenkonzepten zugeordnet werden können.
- Wie in der Versorgungsleitlinie zur Depression (<http://www.versorgungsleitlinien.de>) festgehalten, stellt neben der Diagnose auch der Schweregrad ein wichtiges Merkmal in der Behandlung dar. Dementsprechend könnte auch die mit der Diagnose verbundene Schwere der Erkrankung und damit auch das Profil der Schwere eines Patienten (z.B. Patient mit einer hohen Ausprägung im Problembereich „Aggressivität, Überaktivität“) auch zukünftig als ein Indikator für eine Psychopharmakotherapie betrachtet werden.
- Ein weiteres klinisch hoch relevantes Potential bei der Beurteilung im HoNOS-D stellt die relative Unabhängigkeit von der Berufsgruppe dar. Wie Erfahrungen u.a. aus Großbritannien zeigen, können ebenso Sozialarbeiter, nach eingehender Schulung, die Patienten im HoNOS beurteilen. Damit eröffnet sich eine weitere Perspektive in der klinischen Anwendung. Ebenso wie die Internationale Klassifikation für Funktionsfä-

higkeit, Behinderung und Gesundheit (World Health Organisation, 2001), die die Zielsetzung einer einheitlichen Sprache und damit Verständigung zwischen unterschiedlichen Berufsgruppen verfolgt, könnte auch das patientenspezifische Profil eines Patienten im HoNOS-D u.a. in klinische Fallbesprechungen eingehen und multiprofessionell bezüglich der Therapieplanung diskutiert werden und Berücksichtigung finden.

- Ein wesentliche Möglichkeit zur Verbesserung der klinischen Praxis könnte sich weiterhin daraus ergeben, dass Instrument in die Routine zu implementieren und zum kontinuierlichen Monitoring des Therapieprozesses zu verwenden. In diesem Zusammenhang wäre auch eine Rückmeldung der Einschätzung durch den Therapeuten an den Patienten in verschiedenen Stadien der Therapie denkbar, um somit ablaufende therapeutische Prozesse transparenter zu machen und die Patientenbeteiligung zu stärken.
- Eine weitere klinische Implikation ergibt sich auf Basis der hier vorgestellten empirischen Ergebnisse zur Vorhersage des Ressourcenverbrauches. In unseren Studien erwies sich der Schweregradindex als ein Prädiktor der verbrauchten Ressourcen während stationärer Behandlung. Das gäbe erste Hinweise darauf, dass auch auf der Basis des Schweregrades (in Kombination mit weiteren klinischen Parametern) eine differenzielle Applikation von Ressourcen und eine gezielte Planung des Ressourceneinsatzes möglich wären. Dieses sollte aber nur unter Kontrolle des Therapieergebnisses eingeführt werden.

8.2. Forschungsbezogene Implikationen

Wie eben bereits erwähnt, konnten wir aufzeigen, dass der HoNOS sich als weitest gehend praktikables, reliables und valides Instrument zur differenzierten Beurteilung des Schweregrades (Fremdbeurteilung) erwiesen hat.

Dennoch lassen sich eine Reihe offener Fragen ableiten, womit zukünftiger Forschungsbedarf verbunden wäre:

- So kamen z.B. zur Validierung der HoNOS-D keine spezifischen Instrumente (wie z.B. die BPRS) zur Anwendung. Denkbar wäre eine weiterführende Validierung der

HoNOS-D mit anderen Fremdeinschätzungsinstrumenten, wobei es zurzeit kein vergleichbares Instrument gibt, welches auch für eine größere Gruppe von Patienten relevant wäre.

- Ein besonders hoher Forschungsbedarf ist mit dem weiter entwickelten Item 8 „Andere psychische und verhaltensbezogene Probleme“ verbunden. Hier sind die Entwicklung eines Glossars und damit eine Harmonisierung des Items im Vergleich zu den anderen Items der HoNOS-D dringend angezeigt. Es wäre dann zu prüfen, in wie weit sich die Einschätzung im Item 8, die differenziert und auf der Grundlage von Ankerbeispielen erfolgen würde, zu einer Erhöhung von Reliabilität und Validität dieser wichtigen Modifikation der HoNOS-D beitragen könnte.
- Für mehrere relevante Versorgungsbereiche stehen noch psychometrische Überprüfungen aus (u.a. im ambulanten Bereich und in der psychiatrisch/psychotherapeutischen Behandlung von Patienten mit psychischen Erkrankungen) aus. Auch hier müsste überprüft werden, ob sich der HoNOS-D als ein praktikables, reliables und valides Instrument zur Erfassung des Schweregrades erweisen würde.
- Eine sich weiterhin anschließende Fragestellung betrifft die Untersuchung der differenzierten Erfassung des Schweregrades zu Beginn der Therapie und der Anteil an aufgeklärter Varianz hinsichtlich problematischem interaktionellen Verhaltensweisen und therapeutischer Beziehung im Verlauf der Therapie.
- Ein bereits im vorherigen Abschnitt angesprochene Studie könnte darauf zielen, zu überprüfen, welche Voraussetzungen gegeben sein müssten, damit unabhängige Beurteiler das Therapieergebnis zuverlässig einschätzen könnten und wie sich die Einschätzung im Vergleich zur Beurteilung des Therapeuten unterscheiden würde.

9. Literatur

- Adams, M., Palmer, A., O'Brien, J. T., & Crook, W. (2000). Health of the Nation Outcome Scales for Psychiatry: Are they valid? In *Journal of Mental Health* (Vol. 9, pp. 193-198).
- American Psychiatric Association. (1996). *Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Dt. Bearbeitung und Einführung von Henning Saß*. Göttingen: Hogrefe.
- American Psychiatric Association. (2002). GAF: Skala zur Globalen Erfassung des Funktionsniveaus. In E. Brähler & B. Strauß (Eds.), *Diagnostische Verfahren in der Psychotherapie* (pp. 154-158). Göttingen: Hogrefe-Verlag.
- Amin, S., Singh, S. P., Croudace, T., Jones, P., Medley, I., & Harrison, G. (1999). Evaluating the health of the nation outcome scales: reliability and validity in a three-year follow-up of first-onset psychosis. *The British Journal of Psychiatry*, *174*, 399-403.
- Andreas, S. (2005). *Fallgruppen in der Versorgung von Patienten mit psychischen Störungen - Überprüfung der Eignung eines Fremdeinschätzungsinstrumentes "Die Health of the Nation Outcome Scales, HoNOS-D" zur differenzierten Erfassung des Schweregrades im Rahmen der Entwicklung eines Klassifikationssystems*. Unpublished Dissertation, Universität, Hamburg.
- Andreas, S., Dirmaier, J., Harfst, T., Kawski, S., Koch, U., & Schulz, H. (2009a). On the development and evaluation of a case group concept for inpatients with mental disorders: Using self-report and expert-rated instruments. *European Psychiatry*, *24*, 105-111.
- Andreas, S., Harfst, T., Dirmaier, J., Kawski, S., Koch, U., & Schulz, H. (2007). A psychometric evaluation of the German version of the "Health of the Nation Outcome Scales, HoNOS-D": On the feasibility and reliability of a clinician-rated measure of severity in patients with mental disorders. *Psychopathology*, *40*, 116-125.
- Andreas, S., Harfst, T., Rabung, S., Mestel, R., Schauenburg, H., Hausberg, M., et al. (2010). 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. *International Journal of Methods in Psychiatric Research*, *19*, 50-62.
- Andreas, S., Harris-Hedder, K., Schwenk, W., Koch, U., & Schulz, H. (in press). Is the HoNOS appropriate for the assessment of symptom severity in patients with substance-related disorders? *Journal of Substance Abuse Treatment*.
- Andreas, S., Rabung, S., Horstmann, D., Schauenburg, H., Harris-Hedder, K., Schwenk, W., et al. (2009b). Ist eine Schulung von Ratern zur differenzierten Beurteilung des Schweregrades bei Patienten mit psychischen Störungen in der Routineversorgung notwendig? [Training of mental health care professionals in routine outcome assessment - necessary or dispensable?]. *Psychologische Medizin*, *20*, 117.

- Andreas, S., Rabung, S., Mestel, R., Koch, U., Hausberg, M., & Schulz, H. (submitted). Psychometric evaluation of an extended version of the HoNOS ("Health of the Nation Outcome Scales"): Does the revised item 8 on other mental and behavioral problems work better? *Psychopathology*.
- Andreas, S., Theisen, P., Mestel, R., Koch, U., & Schulz, H. (2009c). Validity of clinical routine DSM-IV diagnoses (Axis I / II) in inpatients with mental disorders. *Psychiatry Research, 170*, 252-255.
- APA. (2000). *Diagnostic Criteria from DSM-IV-TR*. Washington, D.C.: American Psychiatric Association.
- Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP). (1996). AMDP-System. In C. I. P. S. (CIPS) (Ed.), *Internationale Skalen für Psychiatrie* (Vol. 4. überarb. und erw. Aufl.). Göttingen: Beltz-Test.
- Ashcraft, M. L., Fries, B. E., Nerenz, D. R., Falcon, S. P., Srivastava, S. V., Lee, C. Z., et al. (1989). A Psychiatric Patient Classification System: An alternative to Diagnosis-Related Groups. *Medical Care, 27*(5), 543-557.
- Audin, K., Marginson, F. R., Clark, J. M., & Barkham, M. (2001). Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *The British Journal of Psychiatry, 178*, 561-566.
- Baumann, U., & Stieglitz, R.-D. (1989). Evaluation des AMDP-Systems anhand der neueren Literatur (1983-1987) - Überblicksarbeit. *Fortschritte Neurologie Psychiatrie, 57*, 357-373.
- Bech, P., Bille, J., Schutze, T., Sondergaard, S., Wiese, M., & Waarst, S. (2003). Health of the Nation Outcome Scales (HoNOS): Implementability, subscale structure and responsiveness in the daily psychiatric hospital routine over the first 18 months. *Nord J Psychiatry, 57*(4), 285-290.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E., & Erbaugh, J. T. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.
- Bentler, P. M., & Wu, E. (2004). *EQS 6.1 for windows*. Encino CA: Multivariate Software Inc.
- Boot, B., Hall, W., & Andrews, G. (1997). Disability, outcome and case-mix in acute psychiatric in-patient units. *The British Journal of Psychiatry, 171*, 242-246.
- Broadbent, M. (2001). Reconciling the information needs of clinicians, managers and commissioners - a pilot project. *Psychiatric Bulletin, 25*, 423-425.
- Brooks, R. (2000). The reliability and validity of the Health of the Nation Outcome Scales: Validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry, 34*, 504-511.

- Browne, S., Doran, M., & McGauran, S. (2000). Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population. *Irish Journal of Psychological Medicine*, 17(1), 17-19.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park/CA: Sage Publications.
- Buckingham, B., Burgess, P., Solomon, S., Pirkis, J., & Eagar, K. (1998). *Developing a Casemix Classification for Mental Health Services: Summary*. Canberra: Commonwealth Department of Health & Family Services.
- Bundesversicherungsanstalt für Angestellte (BfA). (2000). *Klassifikation therapeutischer Leistungen in der medizinischen Rehabilitation* (4. neubearbeitete ed.). Berlin: BfA.
- Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences* (second ed.). Hillsdale, NJ: Lawrence.
- Collegium Internationale Psychiatriae Salarum (CIPS). (1996a). Brief Psychiatric Rating Scale (BPRS). In Collegium Internationale Psychiatriae Salarum (CIPS) (Ed.), *Internationale Skalen für Psychiatrie* (Vol. 4. überarb. und erw. Aufl., pp. 105-111). Göttingen: Beltz-Test.
- Collegium Internationale Psychiatriae Salarum (CIPS). (1996b). CGI: Clinical Global Impressions. In Collegium Internationale Psychiatriae Salarum (CIPS) (Ed.), *Internationale Skalen für Psychiatrie* (Vol. 4. überarb. und erw. Aufl., pp. 147-151). Göttingen: Beltz-Test.
- Collegium Internationale Psychiatriae Salarum (CIPS). (1996c). Das AMDP-System: Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie. In Collegium Internationale Psychiatriae Salarum (CIPS) (Ed.), *Internationale Skalen für Psychiatrie* (Vol. 4. überarb. und erw. Aufl., pp. 225-233). Göttingen: Beltz-Test.
- Derogatis, L. R. (Ed.). (1977). *SCL-90-R: administration, scoring and procedures manual-I for the r(evised) version*. Baltimore: John Hopkins University School Medicine.
- Eagar, K., Trauer, T., & Mellsoy, G. (2005). Performance of routine outcome measures in adult mental health care. *Australian and New Zealand Journal of Psychiatry*, 39, 713-718.
- Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatry disturbance. *Archives of General Psychiatry*, 33, 766-771.
- Everitt, B. S. (1996). *Making sense of statistics in psychology: A second-level course*. New York: Oxford University Press.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.

- Franke, G. (1995). *SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version*. Göttingen: Beltz.
- Gaines, P., Bower, A., Buckingham, W., Eager, K., Burgess, P., & Green, J. (2003). New Zealand Mental Health Classification and Outcomes Study: Final report. Auckland: Health Research Council of New Zealand.
- Gallagher, J., & Teesson, M. (2000). Measuring disability, need and outcome in Australian community mental health services. *Australian and New Zealand Journal of Psychiatry*, 34, 850-855.
- Goldney, R. D., Fisher, L. J., & Walmsley, S. H. (1998). The Health of the Nation Outcome Scales in psychiatric hospitalisation: A multicentre study examining outcome and prediction of length of stay. *Australian and New Zealand Journal of Psychiatry*, 32, 199-205.
- Harfst, T., Dirmaier, J., Koch, U., & Schulz, H. (2004). Entwicklung und psychometrische Überprüfung einer multidimensionalen Kurzform des Inventars zur Erfassung interpersonaler Probleme. *DRV-Schriften*, 52, 53-55.
- Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D. O., Rüdell, H., & Schulz, H. (2002). Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R [Development and validation of a short version of the Symptom-Check-List-90-R]. *DRV-Schriften*, 33, 71-73.
- Hautzinger, M., Bailer, M., Worrall, H., & Keller, F. (1995). *Das Beck Depressionsinventar (BDI)*. Bern: Huber.
- Hiller, W., Zerssen, D. v., Mombour, W., & Wittchen, H.-U. (1986). *Inpatient Multidimensional Psychiatric Scale. Deutsche Bearbeitung*. Weinheim: Beltz.
- Hirdes, J. P., Fries, B. E., Botz, C., Ensley, C., Marhaba, M., & Pérez, E. (2003). *The System for Classification of In-Patient Psychiatry (SCIPP): A new case-mix methodology for mental health*. Unpublished manuscript, Waterloo ON, Canada.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Hope, J. D., Trauer, T., & Keks, N. A. (1998). Reliability, Validity and utility of the Health of the Nation Outcome Scales (HoNOS) in Australian adult psychiatric services. *Schizophrenia Research*, 29, 9-10.
- Horn, S. D., Chambers, A. F., Sharkey, D. P., & Horn, R. A. (1989). Psychiatric Severity of Illness: A Case Mix Study. *Medical Care*, 27(1), 69-84.
- Horowitz, L. M., Strauß, B., & Kordy, H. (1994). *Inventar zur Erfassung interpersonaler Probleme - Deutsche Version - (IIP-D)*. Weinheim: Beltz.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

- Ilgen, M. A., Quyen, T., Finney, J. W., & Harris, A. H. S. (2005). Recent suicide attempt and the effectiveness of inpatient and outpatient substance use disorder treatment. *Alcoholism, Clinical and Experimental Research*, 29, 1664-1671.
- Issakidis, C., & Teesson, M. (1999). Measurement of need for care: A trial of the Camberwell Assessment of Need and the Health of the Nation Outcome Scales. *Australian and New Zealand Journal of Psychiatry*, 33, 754-759.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, 27(3 Suppl), S178-189.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lange, C., & Heuft, G. (2002). Die Beeinträchtigungsschwere in der psychosomatischen und psychiatrischen Qualitätssicherung: Global Assessment of Functioning Scale (GAF) vs. Beeinträchtigungs-Schwere-Score (BSS). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 48, 256-269.
- Lauzon, S., Corbiere, M., Bonin, J. P., Bonsack, C., Lesage, A. D., & Ricard, N. (2001). [Validation of the French version of the Health of the Nation Outcome Scales (HoNOS-F)]. *Can J Psychiatry*, 46(9), 841-846.
- Linehan, M. (1996). *Dialektisch-Behaviorale Therapie der Borderline-Persönlichkeitsstörung*.
- Lorr, M., Jenkins, R. L., & Holsopple, J. G. (1953). Multidimensional scale for rating psychiatric patients: Hospital form. *Veterans Administration Technical Bulletin*, 43, 10-507.
- Lorr, M., Klett, C. J., McNair, D. M., & Lasky, J. J. (1963). *Manual: Inpatient Multidimensional Psychiatric Scale*. Palo Alto, Cal.: Consulting Psychologists Press.
- McClelland, R., Trimble, P., Fox, M. L., Stevenson, M. R. S., & Bell, B. (2000). Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care*, 9, 98-105.
- Mitchell, J. B., Dicey, B., Liptzin, B., & Sederer, L. J. (1987). Bringing psychiatric patients into the Medicare Prospective Payment System: Alternatives to DRGs. *American Journal of Psychiatry*, 144(5), 610-615.
- National Institute of Mental Health. (1976). 028 CGI. Clinical Global Impressions. In W. Guy & R. R. Bonato (Eds.), *Manual for the EDCEU Assessment Battery* (Vol. 2. rev. Version). Maryland: Chevy Chase.
- Orrell, M., Yard, P., Handysides, J., & Schapira, R. (1999). Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *British Journal of Psychiatry*, 174, 409-412.

- Overall, J. E., & Gorham, D. R. (1976). BPRS. Brief Psychiatric Rating Scale. In W. Guy (Ed.), *ECDEU assessment manual for psychopharmacology* (pp. 157-169). Rockville: National Institute of Mental Health.
- Page, A. C., Hooke, G. R., & Rutherford, E. M. (2001). Measuring mental health outcomes in a private psychiatric clinic: Health of the Nation Outcome Scales and Medical Outcomes Short Form SF-36. *Australian and New Zealand Journal of Psychiatry*, *35*, 377-381.
- Parabiaghi, A., Barbato, A., D'Avanzo, B., Erlicher, A., & Lora, A. (2005). Assessing reliable and clinically significant change on the Health of the Nation Outcome Scales: Method for displaying longitudinal data. *Australian and New Zealand Journal of Psychiatry*, *39*, 719-725.
- Parker, G., O'Donnell, M., Hadzi-Pavlovic, D., & Proberts, M. (2002). Assessing outcome in community mental health patients: A comparative analysis of measures. *International Journal of Social Psychiatry*, *48*(1), 11-19.
- Pirkis, J. E., Burgess, P. M., Kirk, P. K., Dodson, S., Coombs, T., & Williamson, M. K. (2005). A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health and Quality of Life Outcomes*, *3*, 1-12.
- Prinz, U., Nutzinger, D. O., Schulz, H., Petermann, F., Braukhaus, C., & Andreas, S. (2008). Die Symptom-Checkliste-90-R und ihre Kurzversionen: Psychometrische Analysen bei Patienten mit psychischen Erkrankungen. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, *18*, 337-343.
- Rees, A., Richards, A., & Shapiro, D. A. (2004). Utility of the HoNOS in measuring change in a Community Mental Health Care population. *Journal of Mental Health*, *13*, 295-304.
- Rock, D., & Preston, N. (2001). HoNOS: Is there any point in training clinicians? *Journal of Psychiatric and Mental Health Nursing*, *8*, 405-409.
- Salvi, G., Leese, M., & Slade, M. (2005). Routine use of mental health outcome assessments: Choosing the measure. *The British Journal of Psychiatry*, *186*, 146-152.
- Sartorius, N., & Kuyken, W. (1994). Translation of health status instruments. In J. Orley & W. Kuyken (Eds.), *Quality of life assessment: International perspectives* (pp. 3-18). Heidelberg: Springer.
- Saß, H., Wittchen, H.-U., & Zaudig, M. (1998). *Diagnostisches und statistisches Manual psychischer Störungen DSM-IV. Übersetzt nach der vierten Auflage des Diagnostic and Statistical Manual of Mental Disorders der American Psychiatric Association* (3. ed.). Göttingen: Hogrefe.
- Schepank, H. (1995). *Der Beeinträchtigungs-Schwere-Score (BSS)*. Göttingen: Beltz Test.
- Schepank, H. (2002). BSS: Der Beeinträchtigungs-Schwere-Score. In E. Brähler & B. Strauß (Eds.), *Diagnostische Verfahren in der Psychotherapie* (pp. 68-72). Göttingen: Hogrefe-Verlag.

- Schneider, J., Wooff, D., Carpenter, J., Brandon, T., & McNiven, F. (2002). Service organisation, service use and costs of community mental health care. *J Ment Health Policy Econ*, 5(2), 79-87.
- Schulz, H., Harfst, T., Andreas, S., Kowski, S., Koch, U., & Rabung, S. (2009). Zusammenhang zwischen Patienten- und Therapeuteinschätzungen der Ergebnisqualität in der Rehabilitation von Patienten mit psychischen/ psychosomatischen Erkrankungen am Beispiel des SF-8. *Rehabilitation*, 48(270-276).
- Stedman, T., Yellowlees, P., Mellsoy, G., Clarke, R., & Drake, S. (1997). *Measuring consumer outcomes in mental health*. Canberra: ACT: Department of Health and Family Services.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivar Behav Res*, 25, 173-180.
- Stieglitz, R.-D., & Freyberger, H. J. (2001). Selbstbeurteilungsverfahren. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Eds.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (Vol. 2. überarb. und erw. Auflage, pp. 83-92). Stuttgart: Thieme-Verlag.
- Taube, C., Lee, E. S., & Forthofer, R. N. (1984b). DRGs in Psychiatry. An empirical evaluation. *Medical Care*, 22, 597-610.
- Taylor, J. R., & Wilkinson, G. (1997). HoNOS v. GP opinion in a shifted out-patient clinic. *Psychiatric Bulletin*, 21, 483-485.
- Trauer, T. (1999). The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health*, 8, 499-509.
- Trauer, T. (2003). *Analysis of outcome measurement data from the four Victorian "Round One" agencies*. Unpublished manuscript, Melbourne, Australia.
- Trauer, T., Callaly, T., Hantz, P., Little, J., Shields, R., & Smith, J. (1999). Health of the Nation Outcome Scales: Results of the Victorian field trial. *British Journal of Psychiatry*, 174, 380-388.
- Vatnaland, T., Vatnaland, J., Friis, S., & Opjordsmoen, S. (2007). Are GAF scores reliable in routine clinical use? *Acta Psychiatr Scand*, 115(4), 326-330.
- Ware, J. E., Kosinski, M., Dewey, J. E., & Gandek, B. (2000). *How to score and interpret single-item health status measures: A manual for users of the SF-8 Health Survey*: Lincoln: Quality Metric.
- Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B. G., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry*, 172, 11-18.
- Wing, J. K., Curtis, R. H., & Beevor, A. S. (1996). *HoNOS: Health of the Nation Outcome Scales: Report on research and development July 1993-December 1995*. London: Royal College of Psychiatrists.

World Health Organisation. (2001). *The International Classification of Functioning, Disability and Health – ICF*. Geneva.

World Health Organisation (Ed.). (1992). *The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organisation.