

Aus dem Zentralinstitut für Seelische Gesundheit  
der Medizinischen Fakultät Mannheim  
(Direktor: Prof. Dr. med. Andreas Meyer-Lindenberg)

Akzeptanz und Effektivität von computergestütztem prozess- und  
symptombezogenem Therapeutenfeedback im Verlauf stationärer  
Psychotherapie

Inauguraldissertation  
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum)  
der Medizinischen Fakultät Mannheim  
der Ruprecht-Karls-Universität  
zu  
Heidelberg

vorgelegt von  
Christine Gallas

aus  
München  
2019

Dekan: Prof. Dr. med. Sergij Goerd  
Referentin: Prof. (apl.) Dr. sc. hum. Christine Kühner

# INHALTSVERZEICHNIS

Seite

ABKÜRZUNGSVERZEICHNIS .....	6
EINLEITUNG .....	6
1.1 Hinführung zum Thema der Arbeit.....	6
1.2 Das Stuttgart-Heidelberger Modell der Qualitätssicherung.....	7
1.3 Prognose des Therapieerfolgs durch die Behandler.....	9
1.4 Modelle des Gesundheitsverlaufs als Basis für Feedback-Systeme .....	10
1.4.1 Das Aufwand-Wirkungs-Modell .....	11
1.4.2 Implikationen des Aufwand-Wirkungs-Modells .....	13
1.4.3 Das Random Walk Modell .....	13
1.4.4 Implikationen des Random Walk Modells .....	14
1.5 Stand der Forschung zur Effektivität von symptombezogenem Verlaufsfeedback.....	16
1.5.1 Studien der Arbeitsgruppe Lambert und Kollegen .....	16
1.5.2 Deutsche Studien zu Verlaufsfeedback in der Psychotherapie .....	17
1.5.3 Studien anderer Arbeitsgruppen.....	21
1.5.4 Differenzielle Effekte von Verlaufsfeedback in der Psychotherapie ...	22
1.6 Die therapeutische Beziehung.....	26
1.6.1 Das Konzept der therapeutischen Beziehung.....	26
1.6.2 Die Bedeutung der therapeutischen Beziehung für das Therapieergebnis.....	27
1.6.3 Die psychometrische Erfassung der therapeutischen Beziehung .....	29
1.7 Verlaufsfeedback von Prozessmaßen in der Psychotherapie .....	29
1.8 Einordnung der Arbeit in den Forschungsstand .....	30
1.9 Fragestellungen und Hypothesen.....	31
1.9.1 Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen von computergestütztem Therapeutenfeedback .....	32
1.9.2 Fragestellung 2: Gruppenunterschiede bezüglich der Effektivität.....	32
1.9.3 Fragestellung 3: Gruppenunterschiede in der Behandlungsdauer.....	32
1.9.4 Fragestellung 4: Gruppenunterschiede bezüglich der therapeutischen Beziehung .....	33

<b>2 MATERIAL UND METHODEN</b> .....	<b>34</b>
2.1 Rahmenbedingungen und Behandlungssetting.....	34
2.2 Studiendesign.....	34
2.2.1 Ein- und Ausschlusskriterien .....	35
2.3 Studienorganisation und Ablauf der Datenerhebung.....	36
2.4 Erhebungsinstrumente .....	37
2.4.1 Inhaltliche und methodische Aspekte .....	37
2.4.2 Patientenfragebögen .....	38
2.4.3 Therapeutenangaben .....	41
2.5 Inhalte der Rückmeldungen.....	42
2.5.1 Bestimmung von reliabler und klinisch signifikanter Veränderung.....	44
2.5.2 Gruppenunterschiede bezüglich der Rückmeldungen .....	45
2.6 Operationalisierung von Therapieerfolg.....	46
2.6.1 Das Auffälligkeitssignal als aggregiertes Outcome-Maß.....	46
2.7 Auswertung .....	48
<b>3 ERGEBNISSE</b> .....	<b>50</b>
3.1 Stichprobenbeschreibung.....	50
3.1.1 Therapeutenstichprobe.....	50
3.1.2 Patientenfluss .....	50
3.1.3 Soziodemografische Merkmale der Patientenstichprobe.....	51
3.1.4 Klinische Merkmale der Patientenstichprobe.....	53
3.2 Überprüfung der Fragestellungen.....	55
3.2.1 Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen von computergestütztem Therapeutenfeedback .....	55
3.2.2 Fragestellung 2: Vergleich des Therapie-Outcomes.....	60
3.2.3 Fragestellung 3: Vergleich der Behandlungsdauer.....	65
3.2.4 Fragestellung 4: Vergleich der therapeutischen Beziehung.....	68
<b>4 DISKUSSION</b> .....	<b>71</b>
4.1 Repräsentativität der Stichprobe .....	71
4.2 Vergleichbarkeit der Untersuchungsgruppen .....	72
4.3 Zusammenfassung und Diskussion der Ergebnisse (Hypothesenprüfung) .	73

4.3.1	Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen des computergestützten Therapeutenfeedbacks .....	73
4.3.2	Fragestellung 2: Vergleich des Therapie-Outcomes.....	76
4.3.3	Fragestellung 3: Vergleich der Behandlungsdauer.....	78
4.3.4	Fragestellung 4: Vergleich der therapeutischen Beziehung.....	79
4.4	Stärken und Limitationen des Studiendesigns.....	80
4.4.1	Zusätzliches Verlaufsfeedback als „Minimalintervention“ .....	80
4.4.2	Zeitpunkte und Frequenz der Rückmeldungen.....	81
4.4.3	Cluster-Randomisierung .....	81
4.4.4	Auswahl der Erhebungsinstrumente.....	82
4.4.5	Feedback ausschließlich an Einzeltherapeuten.....	83
4.5	Schlussfolgerungen .....	84
4.6	Ausblick .....	85
5	ZUSAMMENFASSUNG.....	87
6	LITERATURVERZEICHNIS.....	89
7	TABELLARISCHER ANHANG.....	103
8	LEBENS LAUF .....	107
9	DANKSAGUNG.....	110

## ABKÜRZUNGSVERZEICHNIS

BSI	Brief Symptom Inventory
BSS	Beeinträchtigungs-Schwere-Score
CST	Clinical Support Tools
GSI	Globaler Schwere-Index
HAQ	Helping Alliance Questionnaire
ICD	International Classification of Diseases
IIP	Inventar zur Erfassung Interpersonaler Probleme
ITT	Intent-to-treat
KI	Konfidenzintervall
NOT	Not on track
OQ	Outcome Questionnaire
OR	Odds ratio
OT	On track
QM	Qualitätsmanagement
QS	Qualitätssicherung
RCI	Reliable Change Index
SCL	Symptomcheckliste
TAU	Treatment as usual
VEV	Veränderungsfragebogen des Erlebens und Verhaltens

## EINLEITUNG

### 1.1 Hinführung zum Thema der Arbeit

Die Wirksamkeit von Psychotherapie wurde durch zahlreiche Studien nachgewiesen und gilt für verschiedene therapeutische Verfahren, Studiendesigns und Settings (zusammenfassend Lambert, 2013). Trotz der insgesamt hohen Effektivität profitiert ein beträchtlicher Anteil an Patienten nicht oder nicht ausreichend von einer Psychotherapie. In Meta-Analysen weisen etwa ein Drittel der Patienten am Ende der Psychotherapie keine signifikante Verbesserung ihres Gesundheitszustands auf (Lambert, 2013). Die ersten Meta-Analysen aus den 80er Jahren berichteten in 10% der untersuchten Fälle negative Effektstärken (Smith, Glass, & Miller, 1980; Shapiro & Shapiro, 1982); aktuell wird von ca. 5% Verschlechterungen ausgegangen (Boisvert & Faust, 2003; Crawford et al., 2016). Während stationärer psychosomatischer Behandlung verschlechtern sich zwischen 5% und 10% der Patienten (Dilcher et al., 2000; Franz et al., 2000). In Settings mit kurzen Behandlungsdauern werden sogar Verschlechterungen von bis zu 23% berichtet und weitere bis zu 40% der Patienten zeigen keine Veränderung (Hansen et al., 2002). Daher besteht auch bei einer insgesamt sehr wirksamen Behandlung wie der Psychotherapie Bedarf nach Methoden, um die Effektivität zu verbessern.

Schwerpunkte der Psychotherapieforschung in den letzten Jahrzehnten waren einerseits kontrollierte Studien mit experimentellem Design und strengen Einschlusskriterien (efficacy-Studien) zur Identifikation von wirksamen *empirically supported treatments* (American Psychological Association, 2006) und zum Anderen Studien, die die Effektivität einer Therapiemethode im naturalistischen Setting untersuchen (effectiveness-Studien). Eine weitere, relativ „junge“ evidenzbasierte Methode ist die patientenzentrierte Forschung (patient-focused research, Howard, Moras, Brill, Martinovich, & Lutz, 1996). Während die efficacy- und effectiveness-Forschungsansätze auf das durchschnittliche Ergebnis einer Therapiemethode für alle untersuchten Patienten fokussieren, versucht der patientenzentrierte Forschungsansatz die Frage „Is this particular treatment working for this patient?“ (Lambert, Hansen, & Finch, 2001, p. 159) zu beantworten. Den Schwerpunkt dieses Ansatzes, dem auch die vorliegende Arbeit zuzuordnen ist, bilden Studien, die den Gesundheitszustand und den Behandlungsverlauf des einzelnen Patienten systematisch erfassen und diese Informationen zeitnah an die Behandler zurückmelden (Lambert, Hansen et al., 2001). Durch das Feedback während der laufenden Psychotherapie soll versucht werden, Misserfolg zu verhindern, die Wahrscheinlichkeit für einen Therapieerfolg zu erhöhen und somit die klinische Praxis zu verbessern. Damit versucht der patientenzentrierte Ansatz, die Kluft zwischen efficacy- und

effectiveness-Forschung auf der einen und der klinischen Praxis auf der anderen Seite – die sog. „researcher-practitioner gap“ (Kazdin, 2008) - zu überbrücken.

Es besteht eine enge Verbindung zwischen patientenzentrierter Forschung und Qualitätssicherung/Qualitätsmanagement. Qualitätsmanagementsysteme wurden ursprünglich für die industrielle Fertigung entwickelt und werden mittlerweile in verschiedenen Adaptationen in der psychotherapeutischen Versorgung angewendet. Im Rahmen des patientenzentrierten Forschungsparadigmas entstanden von Psychotherapieforschern entwickelte Qualitätssicherungssysteme, die individuelle Behandlungsverläufe erfassen und diese Informationen an die Behandler zurückmelden (Barkham et al., 2001; Beutler, 2001; Kordy, Hannover & Richard, 2001; Lambert, Hansen et al., 2001). Diese Systeme unterscheiden sich hinsichtlich ihrer Konzeption und Anwendungsaspekten, teilen jedoch die Zielsetzung, das Behandlungsergebnis des Patienten und damit die Ergebnisqualität zu verbessern (Beutler, 2001).

Im Folgenden wird zunächst das Stuttgart-Heidelberger Modell der Qualitätssicherung dargestellt, auf das sich die Datenerhebung, -auswertung und -rückmeldung in dieser Arbeit stützen. Es schließen sich die Darstellung von Befunden zur Prognose des Therapieerfolgs durch die Behandler an, gefolgt von Modellen des Gesundungsverlaufs in der Psychotherapie, die die Grundlage für Verlaufsfeedback in der Psychotherapie bilden. Ferner wird der Stand der Forschung zur Effektivität von Therapeutenfeedback im Therapieverlauf berichtet, wobei auf differenzielle Feedbackeffekte eingegangen wird. Abschließend werden zentrale Aspekte der therapeutischen Arbeitsbeziehung dargestellt, die in dieser Arbeit als Prozessmaß ausgewählt wurde, um symptomorientiertes Verlaufsfeedback zu ergänzen.

## 1.2 Das Stuttgart-Heidelberger Modell der Qualitätssicherung

In dieser Arbeit wurde eine adaptierte Version der Software Web-AKQUASI, die für das Stuttgart-Heidelberger Modell der QS entwickelt wurde, verwendet. Das Stuttgart-Heidelberger Modell der QS entstand 1992/1993 als Kooperationsprojekt zwischen der Forschungsstelle für Psychotherapie Stuttgart und der Psychiatrischen Universitätsklinik Heidelberg (Lutz et al., 1996; Kordy & Hannover, 1998). Es fokussiert im Vergleich zu anderen, weiter verbreiteten QS-Modellen auf die Behandlungsqualität; Struktur- und Prozessqualität spielen eine untergeordnete Rolle.

Zunächst wurde, basierend auf Erfahrungen zur Psychotherapieevaluation (Kordy, 1992), ein QS-Inventar entwickelt. Dieses sollte ausreichend Informationen liefern, um die Frage „Welche Patienten (Strukturqualität) werden mit welchen therapeutischen Mitteln in welchem Umfang (Prozessqualität) behandelt und welche Ergebnisse werden dabei erreicht (Ergebnisqualität)?“ zu beantworten. Das QS-Inventar erfüllt durch die Berücksichtigung verschiedener Veränderungsbereiche (körperliche, psychische, interpersonale Ebene) und



Bewertungsperspektiven (Patient, Therapeut, bei Bedarf ergänzt durch Pflegeteam u.a.) die Forderung nach einer multidimensionalen und multimodalen Datenerhebung (Schulte, 1993). Ergänzt wurde das Inventar durch den Helping Alliance Questionnaire (HAQ, Bassler, Potratz, & Krauthauser, 1995) zur Erfassung der therapeutischen Arbeitsbeziehung und den ZUF-8 (Schmidt, Lamprecht, & Wittmann, 1989) zur Erhebung der Patientenzufriedenheit, um der im QM geforderten Kundenorientierung („consumer orientation“, World Health Organization, 1997) zu entsprechen.

Im Konsens zwischen Mitgliedern der therapeutischen Teams und Experten der Qualitätssicherung wurde ein Algorithmus zur Bewertung des Therapie-Outcomes entwickelt, der alle Behandlungen nach qualitativen und quantitativen Aspekten in auffällige versus positive Verläufe unterteilt (Abbildung 1). Der Aufmerksamkeitsfokus wird dabei auf die auffälligen Verläufe gelegt. Kriterien für ein Auffälligkeitssignal sind dann erfüllt, wenn entweder ein sog. „Keyitem“ zutrifft (z.B. vom Patienten geäußerte Suizidgedanken oder Verschlechterung aus Therapeutesicht) oder der prä-post-Vergleich der psychometrischen Skalen keine Veränderungen, überwiegend Verschlechterungen oder keine eindeutigen Verbesserungen ergibt.

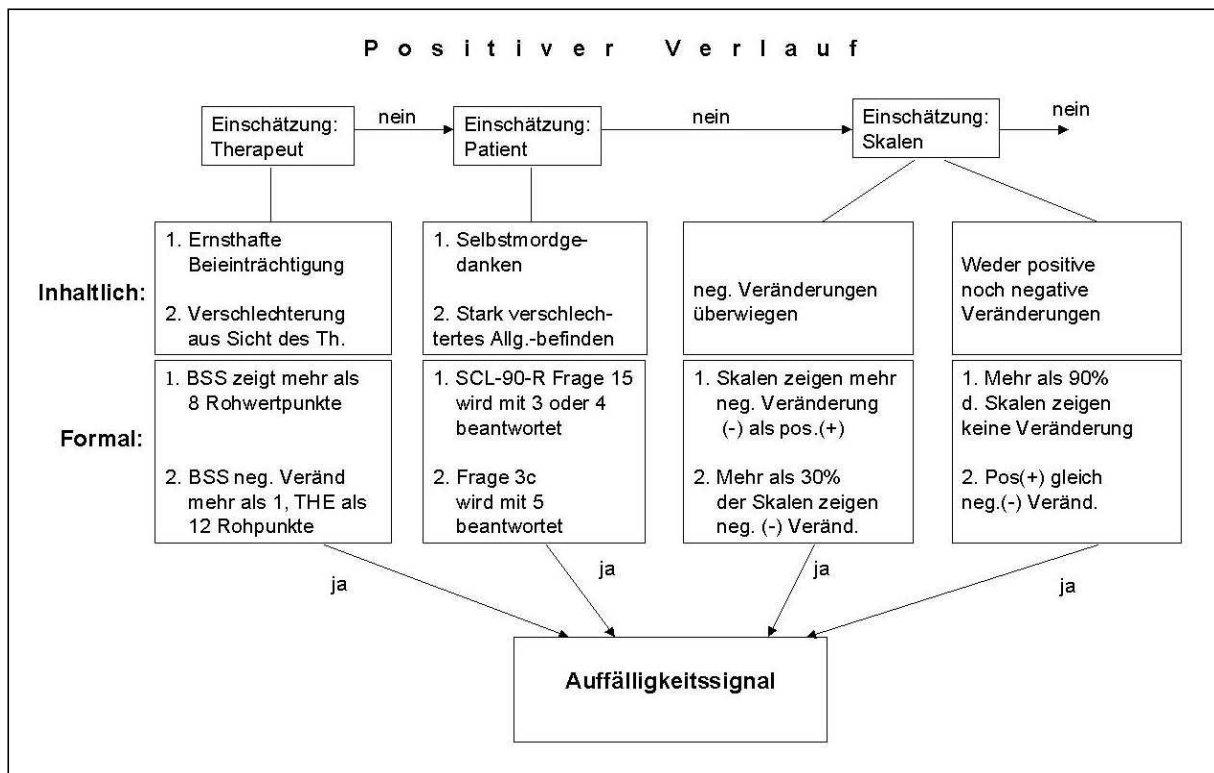


Abbildung 1: Standardisierung der Ergebnisbewertung: Bewertungsalgorithmus nach dem Stuttgart-Heidelberger Modell der QS (nach Kordy & Gallas, 2007, p.986)

Die Validität des Auffälligkeitssignals, das durch a priori festgelegte Kriterien bestimmt wird, wurde anhand von Katamnese-Erhebungen überprüft. In einer Stichprobe von 307 Patienten erwies sich der Gesundheitszustand von bei Entlassung als „positiv“ beurteilten Patienten

auch bei der 6-Monats-Katamnese als sehr verbessert, während sich die Werte für Patienten, die bei Entlassung als „auffällig“ klassifiziert worden waren, bis zur Katamnese sogar weiter verschlechtert hatten (Kordy & Hannover, 1998). Auch in einer größeren Stichprobe von 635 ehemaligen Patienten zeigte sich, dass Personen mit einem zum Entlassungszeitpunkt auffälligen Behandlungsverlauf bei der Nachuntersuchung die ungünstigste Entwicklung aufwiesen (Kordy & Puschner, 2000). In den meisten Fällen resultierte die Bewertung eines Behandlungsverlaufs als „auffällig“ aus Patientenangaben und nicht aus Therapeutenratings (Kordy & Hannover, 1998), was die Bedeutung der Erhebung von Patientendaten in systematisierter Form unterstreicht.

Die Förderung einer Lern- und Kommunikationskultur gilt als zentrales Element eines QS-Systems (Selbmann, 1996). Dafür ist ein effizienter Informationsfluss, der die wesentlichen Informationen zeitnah zur Verfügung stellt, von zentraler Bedeutung. Für das Stuttgart-Heidelberger Modell wurde daher die Software AKQUASI (**AK**tive **QUA**litäts**SI**cherung) entwickelt, die seit den 2000er Jahren in der auf Internettechnologie basierenden Version Web-AKQUASI (Percevic, Gallas, Arikan, Mößner, & Kordy, 2006) vorliegt. Die durch Web-AKQUASI generierten Rückmeldungen werden unter 2.5 dargestellt.

In QS-Systemen stehen Signale, die eine Abweichung vom definierten Zielzustand anzeigen, in der Regel erst nach der Entlassung zur Verfügung und kommen damit für die Behandlung des einzelnen Patienten zu spät. Dieses Manko kann dadurch behoben werden, dass Daten über den Gesundheitszustand des Patienten und einen vom definierten „Soll“ gravierend abweichenden Behandlungsverlauf bereits zeitnah während der Therapie an die Behandler zurückgemeldet werden. Ein solches Verlaufsmonitoring geht über die im QM geforderte Untersuchung der Ergebnisqualität hinaus und bildet einen eigenen Untersuchungsgegenstand im Rahmen des patientenzentrierten Forschungsansatzes.

### 1.3 Prognose des Therapieerfolgs durch die Behandler

Die Rückmeldung von Patientendaten im Therapieverlauf kann nur dann einen Nutzen für die Behandler darstellen, wenn diese Informationen eine sinnvolle Ergänzung der eigenen Einschätzung darstellen. Die Validität der Beurteilung des Therapieverlaufs und des prognostizierten Therapieerfolgs durch Therapeuten wurde mehrfach untersucht. Dabei zeigte sich, dass das klinische Urteil alleine nicht geeignet ist, Therapieverlauf, Therapieerfolg oder Misserfolg vorherzusagen (Grove, 2005; Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010; Meyer & Schulte, 2002; Schulte et al., 1991).

In einer Studie von Hannan et al. (2005) sollten Therapeuten einschätzen, bei welchen Patienten sich die Symptomatik verschlechtern wird, wobei nur einer von 40 entsprechenden Fällen von den Behandlern korrekt identifiziert werden konnte. Hatfield et al. (2010) verglichen

die Verlaufsdocumentationen, die Therapeuten von ihren Sitzungen anfertigten, mit der Selbsteinschätzung von Patienten auf Symptomskalen. Dabei zeigte sich, dass nur in 32% der Fälle, in denen sich die Symptomatik innerhalb der Woche vor der Therapiesitzung deutlich verschlechtert hatte, die Therapeuten eine Verschlechterung des Befindens notiert hatten.

Der Befund, dass Therapeuten die Wahrscheinlichkeit für einen Therapieerfolg überschätzen, zeigt sich seit mehreren Jahrzehnten stabil (z.B. Bergin, 1971; Hannan et al., 2005; Walfish, McAlister, O'Donnell, & Lambert, 2012). Dieser sog. „self-assessment bias“ – eigene Leistung und Fähigkeiten als überdurchschnittlich im Vergleich zu Anderen einzuschätzen - ist allerdings kein für die psychotherapeutische Behandlung spezifisches Phänomen, sondern zeigt sich auch in vielen anderen Lebensbereichen (z.B. Dunning, Heath, & Suls, 2004).

Eine Studie an einer universitären Psychotherapie-Ambulanz zeigte, dass die Erfolgsbeurteilungen einzelner Sitzungen und die Prognose des Therapieerfolgs durch die Therapeuten nicht oder allenfalls gering mit dem späteren Behandlungserfolg korrelierten (Meyer & Schulte, 2002). Dabei spielte es kaum eine Rolle, ob der Behandlungserfolg aus Sicht der Therapeuten selbst, aus Patientensicht oder über den prä-post-Vergleich psychometrischer Skalen eingeschätzt wurde. Auch wenn Erfolgsprognose und späterer Behandlungserfolg beide aus Therapeutensicht beurteilt worden waren, betrug der Zusammenhang nur maximal  $r=.36$ . Auch in einer früheren Studie, in der ausschließlich Therapeutenratings verwendet wurden, korrelierte die Erfolgseinschätzung der einzelnen Sitzungen und die Erfolgsprognose nur gering mit dem Therapieerfolg ( $r=.20$ , Schulte, Künzel, Pepping, & Schulte-Bahrenberg, 1991). Die Validität des klinischen Urteils scheint außerdem nicht mit der Berufserfahrung des Therapeuten zusammen zu hängen (Dawes, 1989; Grove & Meehl, 1996; Hannan et al., 2005).

Obwohl die Validität der Erfolgsprognose durch Therapeuten als eher gering einzuschätzen ist, waren sich die Therapeuten hinsichtlich des Zutreffens ihrer Prognose relativ sicher (Meyer & Schulte, 2002). Schulte und Meyer (2002) untersuchten, woran Therapeuten sich bei der Erfolgsprognose orientieren. Es zeigte sich, dass die Erfolgsprognose überwiegend durch das Ausmaß der positiven Stimmung geprägt wird, die wiederum davon abhängt, ob sich der Therapeut von seinem Patienten akzeptiert oder abgelehnt fühlt. Daher „...erscheint es ratsam, dass die Therapeuten ihre Entscheidungen über ihr Vorgehen an objektiveren Daten ausrichten...“ (Meyer & Schulte, 2002, p.60).

#### 1.4 Modelle des Gesundungsverlaufs als Basis für Feedback-Systeme

Daten, die an die Behandler über den Gesundheitszustand und den Therapieverlauf ihrer Patienten zurückgemeldet werden, müssen interpretiert werden, damit sie für die Therapieplanung genutzt werden können. Die Bewertung kann aus klinischer Sicht unterschiedlich ausfallen. Ein weit über der Norm liegender Wert einer Symptomskala wird

üblicherweise als Ausdruck einer starken Beeinträchtigung bewertet. In Einzelfällen kann dies aber auch anders gesehen werden – z.B. als Verbesserung des Problembewusstseins bei Patienten, die zuvor zur Bagatellisierung geneigt hatten. Auch eine rasche Veränderung auf Symptomskalen kann unterschiedlich interpretiert werden, z.B. als eine erfreulich schnelle Besserung des Gesundheitszustands oder als „sozial erwünschtes“ Antwortverhalten. Auch stellt sich die Frage, wie mit einer ausbleibenden Veränderung umgegangen werden soll: Ist dies als Hinweis zu sehen, dass die bisherige Behandlungsstrategie zu ändern ist? Benötigt der Patient eine andere Therapieform oder sollte gar ein Therapeutenwechsel in Betracht gezogen werden? Oder ist eine Änderung des Behandlungsplans eine vorschnelle Maßnahme, die unnötige Störungen im Therapieverlauf hervorruft?

Welche Interpretationshilfen Behandler seitens der Psychotherapieforschung erhalten, hängt davon ab, welche Modelle oder empirischen Befunde zugrunde gelegt werden. Im Folgenden werden zwei Modelle, auf die sich aktuelle Monitoring- und Feedbacksysteme beziehen, beschrieben: Das Aufwand-Wirkungs-Modell sowie das Random Walk Modell.

#### 1.4.1 Das Aufwand-Wirkungs-Modell

Die meisten in der Psychotherapie verwendeten Monitoring- und Feedbacksysteme leiten ihre Feedbackstrategie aus dem Aufwand-Wirkungs-Modell (dose-response model) ab sowie Befunden zur Bedeutung einer „early response“ (Symptomverbesserung zu einem frühen Zeitpunkt im Therapieverlauf). Das Aufwand-Wirkungs-Modell überträgt die aus der Medizin stammende Frage „Welche Menge eines Wirkstoffs ist notwendig, um welchen Effekt zu erzielen?“ auf die Psychotherapie. Therapiedauer und Sitzungszahl („Dosis“) werden mit dem Ergebnis („Wirkung“) – in der Regel operationalisiert durch die prä-post-Veränderung auf Symptomskalen – in Beziehung gesetzt. Howard, Kopta, Krause und Orlinsky (1986) fanden in einer Meta-Analyse, die 15 Studien mit mehr als 2000 Patienten umfasste, dass sich die Beziehung zwischen der Anzahl an Therapiesitzungen und der Verbesserungsrate log-linear, als negativ beschleunigte Kurve mit positiver Steigung, darstellte. Sie stellten fest, dass sich bei 41% der Patienten bereits nach der vierten Sitzung und bei insgesamt 53% nach der achten Sitzung eine deutliche Verbesserung der Symptomatik zeigte. Um eine weitere Steigerung der Verbesserungsrate zu erreichen, war ein zunehmend größerer Therapieaufwand nötig. 75% der Patienten erreichten eine Verbesserung bis zur 26. Sitzung und bei etwa 60 Sitzungen erreicht die Verbesserungsrate mit 85% die Asymptote (Howard et al., 1986). Dies wird auch als „law of diminishing returns“ bezeichnet und als Begründung herangezogen, um die Allokation psychotherapeutischer Ressourcen aus Effizienzgründen grundsätzlich zu beschränken: „In clinics that serve a large population with limited resources, 26 sessions might be used as a rationale time limit“, so die Schlussfolgerung der Autoren (Howard et al., 1986, p. 163). Kopta, Howard, Lowry und Beutler (1994) berichten, dass die

benötigte Dosis je nach verwendetem Ergebnismaß unterschiedlich ausfällt; dennoch beschreiben sie die negativ beschleunigte Kurve als gemeinsames Muster der Veränderungsraten auf verschiedenen Ergebnisdimensionen. Die über verschiedene Stichproben aggregierten Dosis-Wirkungs-Kurven werden meist so verstanden, dass sie den Durchschnitt einzelner, ebenfalls negativ beschleunigter Dosis-Wirkungs-Kurven darstellen. Dies ist jedoch nicht notwendigerweise der Fall, da sich die Patienten in den einzelnen Studien, die von Howard et al. (1986) und Kopta et al. (1994) untersucht und zusammengefasst wurden, unterschiedlich lange in Therapie befanden. Damit finden sich zu jedem berichteten Zeitpunkt unterschiedlich aggregierte Stichproben und der aggregierte Schätzwert, welchen Zugewinn an Effektivität jede zusätzliche Sitzung bringt, basiert mit zunehmender Sitzungszahl auf immer kleiner werdenden Subgruppen aus „Slow-Respondern“ (Patienten, die nur langsam auf die Therapie ansprechen) und „Non-Respondern“ (Patienten, die gar nicht auf die Therapie ansprechen). Barkham et al. (1996) argumentieren, dass die einzelnen Dosis-Wirkungs-Kurven zwar unterschiedliche Steigungen aufweisen, alle aber eine lineare statt eine logarithmische Form haben könnten. Dies wäre z.B. dann der Fall, wenn die Verbesserung der Problematik über die einzelnen Sitzungen hinweg kontinuierlich (linear) erfolgen würde, bis ein Punkt erreicht ist, an dem Patient und Therapeut sich entweder anderen Problemen zuwenden oder aufgrund der Zielerreichung die Therapie beenden. Würden diese einzelnen linearen Kurven dann aggregiert, hätte die durchschnittliche Dosis-Wirkungs-Kurve eine logarithmische Form, da Patienten mit Problemkonstellationen, bei denen schnelle Verbesserungen erzielt werden, die Behandlung früher beenden.

Ausgehend von einer negativ beschleunigten Aufwand-Wirkungs-Kurve, die eine höhere Besserungsrate in den ersten gegenüber späteren Therapiesitzungen beschreibt, wurden eine Reihe von Studien durchgeführt, die auf individueller Ebene die Bedeutung einer frühen Symptomverbesserung („early response“) für den Therapieerfolg untersuchten. In den meisten dieser Arbeiten wurde das Ausmaß an früher Symptomveränderung mit dem Therapieergebnis – d.h., der Gesamtveränderung - in Beziehung gesetzt, wobei sich die frühe Änderung als prädiktiv für die gesamte Veränderung erwies. Beispielsweise zeigten Brown und Lambert (1998) anhand großer Datensätze, dass die initiale Beeinträchtigungsschwere und die Veränderung nach der ersten Therapiesitzung 17% der Varianz des Therapieergebnisses erklärten. Wurde die Veränderung bis zur dritten Sitzung in die Analyse einbezogen, konnte sogar 40% der Varianz des Therapieerfolgs erklärt werden. Der Befund, dass eine frühe Therapieresponse erfolgreich behandelte von nicht erfolgreich behandelten Patienten unterscheidet, konnte in verschiedenen Studien in unterschiedlichen Settings repliziert werden (z.B. deJong-Meyer, Hautzinger, Rudolf, Strauß, & Frick, 1996; Haas, Hill, Lambert, & Morrell, 2002; Stiles et al., 2003; Tang & DeRubeis, 1999).

### 1.4.2 Implikationen des Aufwand-Wirkungs-Modells

Basierend auf dem Aufwand-Wirkungs-Modell und Befunden zur „early response“ kommt einem Monitoring- und Feedbacksystem die zentrale Aufgabe zu, Indikatoren für „Response“ bzw. „Non-Response“ zu erfassen und zeitnah zurück zu melden, damit die laufende Therapie entsprechend modifiziert werden kann (Howard et al., 1996). In den Monitoring- und Feedbacksystemen, die sich auf das Aufwand-Wirkungs-Modell stützen, wird der Therapieverlauf des einzelnen Patienten mit dem erwarteten (modellierten) Verlauf, basierend auf den Daten einer Vergleichspopulation verglichen (Finch, Lambert & Schaalje, 2001; Lambert, Hansen et al., 2001; Lueger et al. 2001; Lutz, Martinovich, & Howard, 1999). Als erwarteter Verlauf wird dabei in der Regel eine Symptomverbesserung definiert, die der durchschnittlichen Veränderungsgeschwindigkeit der Vergleichsgruppe (innerhalb eines Konfidenzintervalls) entspricht. Das Ansprechen auf die Behandlung im frühen Verlauf wird dabei als prädiktiv für den weiteren therapeutischen Fortschritt angenommen. Ausbleibende Symptomverbesserungen im frühen Therapieverlauf werden als Prädiktor für therapeutischen Misserfolg und damit als Alarmsignal gewertet. Demnach empfehlen Feedbacksysteme, die sich auf das Aufwand-Wirkungs-Modell stützen, bei einer „early non-response“ die Gründe für das Nichtansprechen des Patienten auf die Therapie zu untersuchen (z.B. mangelnde Therapiemotivation, schlechte therapeutische Beziehung usw.), den therapeutischen Ansatz zu überdenken, die therapeutischen Maßnahmen zu ändern oder sogar die Therapiemethode oder den Therapeuten zu wechseln. Studien zur Effektivität von Verlaufsfeedback an Therapeuten, das auf den Entscheidungsregeln solcher Vorhersagemodelle beruht, werden in Kapitel 1.5.1 dargestellt.

### 1.4.3 Das Random Walk Modell

Feedbacksystemen, die sich auf das Dosis-Wirkungs-Modell und die Bedeutung einer „early response“ stützen, liegt die Annahme zugrunde, dass vom Symptomverlauf in der Vergangenheit („response“) auf den zukünftigen Gesundungsverlauf geschlossen werden kann, so dass bei Ausbleiben des erwarteten Ansprechens auf die Behandlung ein Alarmsignal geben wird. Allerdings wurde die Annahme, dass der Gesundungsverlauf in der Vergangenheit prädiktiv ist für zukünftige Veränderungen, in den Studien zur Bedeutung einer „early response“ nicht empirisch getestet. Untersucht wurde vielmehr der Zusammenhang zwischen einer substanziellen Symptomverbesserung in einer frühen Phase der Therapie und dem Therapieerfolg, der üblicherweise als die Gesamtveränderung zwischen Behandlungsbeginn und -ende operationalisiert wird. Da die frühe Veränderung jedoch immer Teil der Gesamtveränderung ist, interkorrelieren „early response“ und Therapieerfolg notwendigerweise. Daher wurde vorgeschlagen, den Einfluss von Symptomveränderungen auf *nachfolgende Änderungsraten* – und nicht auf das Therapieergebnis – zu untersuchen

(Percevic, Lambert, & Kordy, 2004, 2006). Percevic et al. (2006) untersuchten Symptomverläufe und die Abhängigkeit aufeinanderfolgender Veränderungsraten anhand dreier Stichproben aus Deutschland und den USA (ambulantes und stationäres Setting). Dabei ließen sich die Symptomverläufe sowohl durch ein lineares als auch ein logarithmisches Modell darstellen, wobei in beiden Fällen der Vorhersagefehler beträchtlich war. Gleichzeitig zeigten sich aufeinanderfolgende Veränderungsraten als voneinander unabhängig (Percevic et al., 2006). Bei unabhängigen Veränderungsraten ist die Wahrscheinlichkeit für eine Veränderung nicht vom Zeitpunkt im Therapieverlauf abhängig oder vom Ausmaß der bereits eingetretenen Veränderung. Da nichtdeterministische Zeitreihen auch als „random walks“ bezeichnet werden, wurde dieses Modell zur Beschreibung von Gesundungsverläufen in der Psychotherapie „Random Walk Modell“ genannt (Percevic et al., 2004).

Neben den bereits beschriebenen methodischen Problemen bei Howard et al. (1986) ist ebenfalls kritisch zu werten, dass Studien, die eine negativ beschleunigte (logarithmische) Funktion als geeignet für die Darstellung von Therapieverläufen beschreiben, nicht berichten, ob alternative (insb. einfachere lineare) Verlaufsmodelle getestet worden waren (Lambert, Hansen, et al., 2001; Lutz, Martinovich, et al., 2001). Zudem existieren neben den Befunden von Percevic et al. (2004, 2006) weitere Studien, in denen sich Gesundungsverläufe durch lineare Funktionen (mit hoher interindividueller Variation) darstellen ließen (Barkham, Stiles, & Shapiro, 1993; Barkham et al., 1996; Puschner, Kraft, Kächele, & Kordy, 2007; McClintock, Perlman, McCarrick, Anderson, & Himawan, 2017). Bei der Untersuchung der Dosis-Wirkungs-Beziehung auf der Grundlage wöchentlicher Symptomratings fanden Kadera, Lambert und Andrews (1996) eine hohe Variabilität bezüglich der individuellen Veränderungsmuster, mit einem allgemeinen linearen Trend. Daten aus der ambulanten Richtlinienpsychotherapie (Deutschland) mit gegenüber den USA oder Großbritannien deutlich längerer Therapiedauer zeigten ebenfalls keine Überlegenheit des logarithmischen Modells gegenüber einem linearen Modell bei der Schätzung des Symptomverlaufs, so dass dem einfacheren (linearen) Modell der Vorzug gegeben wurde (Puschner et al., 2007).

#### 1.4.4 Implikationen des Random Walk Modells

Der Befund voneinander unabhängiger Änderungsraten unterstützt nicht den Ansatz einer „expected treatment response“, der in Feedbacksystemen, die auf dem Aufwand-Wirkungs-Modell basieren, verfolgt wird und bei (früher) Non-Response ein Alarmsignal wegen drohendem Misserfolg der Behandlung sendet (Howard et al., 1996; Lutz et al., 1999). Vielmehr spricht die Unabhängigkeit von aufeinander folgenden Änderungsraten dafür, dass ein Patient unabhängig von bereits stattgefundenen Veränderung die Chance hat, von einer Fortsetzung der Therapie zu profitieren. Dies bedeutet für ein Monitoring- und Feedbacksystem, das sich auf das „Random Walk Modell“ stützt, im Falle einer frühen Non-

Response *kein* Alarmsignal zu geben. Da ein „Non-Responder“ bei einer streng limitierten Behandlungsdauer allerdings eine geringere Chance auf ein positives Therapieergebnis hat als ein „early Responder“, sollte das Therapiekontingent flexibel sein, um ausreichend Zeit für Veränderungen zu schaffen. Brown und Jones (2005) beantworten die Frage „Implementation of a feedback system in a managed care environment: What are patients teaching us?“ wie folgt: „The most important message that emerges from these data is that patients who have high levels of distress have a high probability of improvement, providing they remain engaged in treatment for a sufficient period.“ (pp. 196-197). Dabei betonen die Autoren die Bedeutung, die einer optimalen Nutzung begrenzter Ressourcen durch die Anpassung von Behandlungsdauer und -intensität an die individuellen Bedürfnisse des Patienten zukommt. Da Patienten sehr unterschiedlich auf Psychotherapie ansprechen und unabhängige Veränderungsraten angenommen werden, erscheint bei Zugrundelegung des „Random Walk Modells“ eine a priori Kontingentierung von Therapiedauer oder Sitzungszahl nicht angemessen. Die Erwartung, bei Patienten mit ausbleibender oder geringer Response durch eine Verlängerung der Therapie eine Symptomverbesserung zu erreichen, scheint allerdings dem oben beschriebenen Dosis-Wirkungs-Modell zu widersprechen. Da beide dargestellte Modelle nur aggregierte Schätzwerte des Veränderungsverlaufs liefern – mit großer interindividueller Varianz bezüglich der Veränderungsgeschwindigkeit und des Therapieverlaufs – sollten die Ergebnisse modellierter, aggregierter Daten nicht alleinige Grundlage sein, um den individuellen Gesundheitsverlauf des einzelnen Patienten zu monitorieren. Vielmehr scheint es sinnvoll, im Sinne des „good enough level“-Modells (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009) anzunehmen, dass Patient (und Therapeut) die Therapie beenden, wenn die individuellen Therapieziele erreicht sind oder zufriedenstellende Verbesserungen eingetreten sind. Die Aufgabe eines Feedbacksystems besteht darin, den Grad der Therapiezielerreichung und das Ausmaß an Veränderungen im Therapieverlauf zu monitorieren, um die Entscheidung über Therapiefortsetzung oder –beendigung empirisch zu unterstützen (vgl. Abbildung 2).

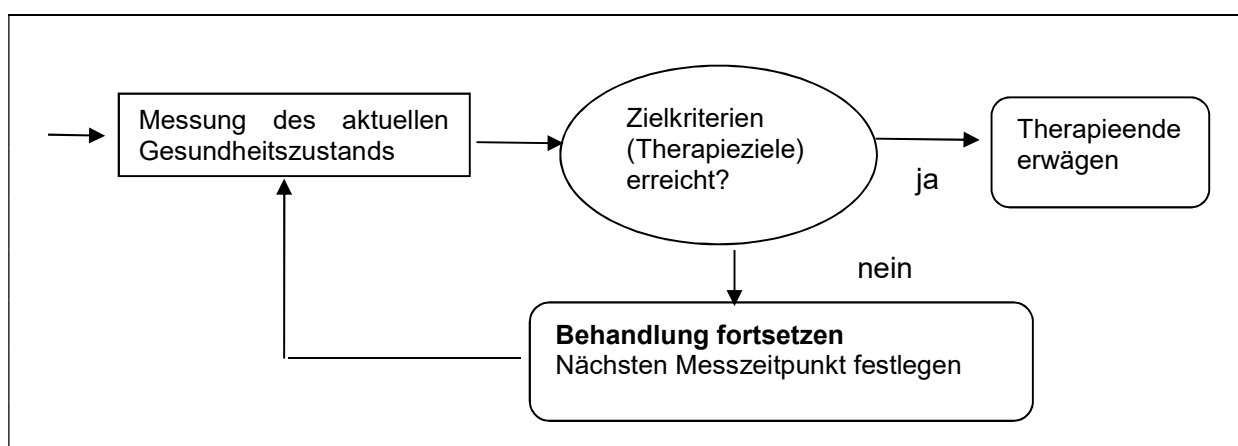


Abbildung 2: Monitoring-Strategie basierend auf dem Random Walk Modell mit der Annahme voneinander unabhängigen Veränderungsraten (nach Percevic et al., 2006)



## 1.5 Stand der Forschung zur Effektivität von symptombezogenem Verlaufsfeedback

### 1.5.1 Studien der Arbeitsgruppe Lambert und Kollegen

Die „Pionierarbeit“ bezüglich der Rückmeldung psychometrischer Verlaufsdaten aus Patientensicht an die behandelnden Therapeuten leistete die Arbeitsgruppe um M. J. Lambert (Utah, USA). Den theoretischen Hintergrund des Feedbacksystems bildet das oben dargestellte Dosis-Wirkungs-Modell; einer „early response“ wird dabei große Bedeutung beigemessen (Lambert, Whipple et al., 2001). Der Gesundungsverlauf des Patienten wird mit dem einer klinischen Vergleichspopulation verglichen, wobei die Beeinträchtigung bei Therapiebeginn, die Anzahl absolvierter Therapiestunden und das bereits erreichte Ausmaß an Veränderung berücksichtigt werden. Entsprechend wird der Behandlungsverlauf als „on track“ (OT, Veränderung des Patienten ist angemessen) oder „not on track“ (NOT, Patient zeigt nicht den erwarteten Fortschritt) bewertet (Lambert, Whipple, Bishop et al., 2002). Die Symptomatik wird durch den Outcome Questionnaire-45 (OQ-45, Lambert et al., 1996; Lambert et al., 2013) erfasst, der wöchentlich vor den Sitzungen erhoben wird. Die Feedbacknachricht besteht aus einem Verlaufsgraphen, der den OQ-45-Gesamtscore zu den verschiedenen Messzeitpunkten darstellt, sowie einer farblichen Markierung, die bei einem „not on track“-Fall Signalcharakter hat (gelb: Veränderungsrate des Patienten ist unterdurchschnittlich, rot: Patient zeigt keine Veränderung oder Symptomatik hat sich verschlechtert). Bei „not on track“-Fällen erhalten Therapeuten die Rückmeldung, dass die Wahrscheinlichkeit eines ausbleibenden Therapieerfolgs hoch ist und eine Änderung des Behandlungsplans in Betracht gezogen werden sollte. Die Arbeitsgruppe legte Ergebnisse aus sechs Studien vor (Harmon, Hawkins, Lambert, Slade, & Whipple, 2005; Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Slade, Lambert, Harmon, Smart, & Bailey, 2008; Lambert, Whipple et al., 2001; Lambert, Whipple, Vermeersch et al., 2002; Whipple et al., 2003). Alle Studien wurden im ambulanten Setting durchgeführt; mit Ausnahme einer Studie (Hawkins et al., 2004: Klinikambulanz) alle an einer Beratungsstelle für Studierende. Zusammenfassend zeigen die Studien bessere Behandlungsergebnisse für Patienten, deren Therapeuten Verlaufsfeedback erhalten hatten (Metaanalyse der sechs Studien mit insgesamt 6.151 Patienten, Shimokawa, Lambert, & Smart, 2010). Die durchschnittliche Effektstärke (OQ-45 Gesamtwert bei Therapieende) betrug über alle sechs Studien hinweg  $g = -0.12$ . Lambert et al. (2003) fanden in einer ersten Meta-Analyse ihrer eigenen Studien, dass dieser Effekt auf Rückmeldungen bei NOT-Patienten (durchschnittliche Effektstärke 0.39) zurück zu führen ist – Rückmeldungen bei OT-Patienten zeigten lediglich eine Effektstärke von 0.09. In der Meta-Analyse von Shimokawa et al. (2010) wird bei NOT-Fällen eine Effektstärke von  $g = -0.28$  zugunsten der Feedbackbedingung berichtet.

Weitere Studien zielten darauf ab, den Feedbackeffekt zu stärken, indem zusätzlich Rückmeldungen an Patienten gegeben wurden und/oder die Rückmeldungen durch sog.

„clinical support tools“ (CST) ergänzt wurden. In der CST-Bedingung werden – ausschließlich bei NOT-Patienten - Maße zur therapeutischen Beziehung (HAQ-II, Luborsky et al., 1996), zur Änderungsmotivation (McConaughy, Prochaska, & Velicer, 1983) und zur sozialen Unterstützung erhoben und an die Therapeuten zurückgemeldet, kombiniert mit Empfehlungen für therapeutische Maßnahmen. Die Auswahl der Konstrukte orientierte sich an der Literatur zur Prozess-Ergebnis-Forschung in der Psychotherapie (Whipple et al., 2003). Whipple et al. (2003) berichten für Patienten, deren Therapeuten die „clinical support tools“ angewandt hatten, signifikant niedrigere Symptomwerte bei Behandlungsende im Vergleich zur „herkömmlichen“ Feedbackintervention, die wiederum TAU (kein Feedback) leicht überlegen war. Die Studie von Hawkins et al. (2004) ergab in der Gesamtstichprobe einen kleinen zusätzlichen positiven Effekt, wenn Feedback nicht allein an Therapeuten, sondern auch an Patienten gegeben wurde. In einer weiteren, randomisierten Studie (Harmon et al., 2007) konnte der zusätzliche positive Effekt der CST-Intervention auf das Behandlungsergebnis repliziert werden, allerdings wurde im Gegensatz zu Hawkins et al. (2004) kein Effekt des zusätzlichen Patientenfeedbacks gefunden. Bei Slade et al. (2008) wurde ebenfalls kein zusätzlicher Effekt von Feedback an Patienten gefunden, jedoch ein positiver Zusatzeffekt, wenn das Feedback um die „clinical support tools“ ergänzt wurde.

## 1.5.2 Deutsche Studien zu Verlaufsfeedback in der Psychotherapie

### 1.5.2.1 Stationäres Setting

In der ersten in Deutschland durchgeführten Studie zu Verlaufsfeedback in der Psychotherapie untersuchte Bauer (2004) in einer prospektiven Studie mit sequenziellem Design die Effektivität einer einmaligen ergebnisbezogenen Rückmeldung (vier Wochen nach Aufnahme) im Verlauf stationärer Langzeitpsychotherapie. Dabei stützte sich die Rückmeldung in der ersten Phase der Studie (n=109) auf das Konzept der „early treatment response“ (vgl. 1.4.1). In der zweiten Studienphase an 104 Patienten wurde zusätzlich eine Prognose der Behandlungsdauer und eine Behandlungsempfehlung in Anlehnung an das Random Walk Modell (vgl. 1.4.3) gegeben. Die Daten wurden in Papier-Bleistift-Form erhoben, eingescannt, ausgewertet und wenige Tage später wurde eine schriftliche Rückmeldung an die Behandler ausgehändigt. In beiden Phasen zeigte sich kein Effekt der Feedbackintervention auf das Behandlungsergebnis. Allerdings wurde ausschließlich in der zweiten Phase, in der die Rückmeldung eine Empfehlung hinsichtlich der Fortführung oder Beendigung der Behandlung enthielt, ein vergleichbares Behandlungsergebnis bei kürzerer Therapiedauer erreicht, so dass diese Feedbackstrategie mit einer höheren Effizienz verbunden war. Kritisch ist zu sehen, dass die Gruppenzuteilung nicht randomisiert erfolgte; die Vergleichsgruppen wurden aus Patienten gebildet, bei denen aus organisatorischen Gründen kein Therapeutenfeedback möglich war.

In einer weiteren Studie wurden 118 Patienten einer Rehabilitationsfachklinik für Psychosomatik, Psychotherapie und Verhaltensmedizin in zwei Bedingungen mit und ohne Verlaufsfeedback randomisiert (Berking, Orth, & Lutz, 2006). Die psychische Beeinträchtigung aus Patientensicht wurde durch Kurzformen gängiger psychometrischer Instrumente erfasst, die zu einem Gesamtmaß (FEP: Fragebögen zur Erfassung von Erfolgen und Verläufen psychotherapeutischer Behandlungen, Lutz, Tholen, Schürch, & Berking, 2006) verrechnet wurden. Datenerhebung und Feedback an die Therapeuten erfolgten in Papierform bei Aufnahme sowie im wöchentlichen Rhythmus bis zur Entlassung. Zurück gemeldet wurde die aktuelle Symptomatik des Patienten; auf die Darstellung von Veränderungsgraphen oder Algorithmen zur Bewertung des Behandlungsverlaufs (Signale) wurde verzichtet. Es zeigten sich auf allen Erfolgsmaßen größere Effekte bei Patienten in der Feedbackgruppe als bei Patienten der Kontrollgruppe ( $d=0.47 - 0.50$ ). Die Autoren vermuten als Wirkmechanismus des Feedbacks eine erhöhte Aufmerksamkeit und verstärktes Engagement des Behandlerteams. Die im Vergleich zu den Studien der Lambert-Gruppe großen Feedbackeffekte – lediglich in der Untersuchung von Hawkins et al. (2004) mit stärker beeinträchtigten Patienten einer Krankenhausambulanz wurde mit  $d=0.17$  ein signifikanter Effekt für die Gesamtgruppe berichtet – erklären die Autoren mit größeren therapeutischen Einflussmöglichkeiten im stationären Setting.

Eine weitere Untersuchung zur Effektivität von Verlaufsfeedback wurde in der stationären Psychiatrie durchgeführt (Puschner, Schöfer, Knaup, & Becker, 2009). In einem cluster-randomisierten Design wurden Therapeuten einer Feedbackbedingung (Patienten- und Therapeutenfeedback) oder TAU (kein Feedback) zugewiesen. In beiden Bedingungen beantworteten die Patienten zum Aufnahmezeitpunkt, jede Woche während der Behandlung sowie bei Entlassung am Computer die deutsche Version des OQ-45, den Ergebnisfragebogen EB-45 (Lambert, Hannover, Nisslmüller, Richard, & Kordy, 2002). In der Feedbackbedingung erhielten die Therapeuten wöchentlich eine schriftliche Rückmeldung, die die Verlaufskurve der OQ-45-Gesamtwerte, farbliche Signale (wie im Feedbacksystem von Lambert und Kollegen), eine schriftliche Zusammenfassung von Symptomstatus und -verlauf, Informationen über Suizidalität sowie eine Empfehlung bezüglich der Fortsetzung der Behandlung beinhaltete. Letztere stützte sich auf das Random Walk Modell: Befand sich der OQ-45-Gesamtwert zum letzten Messzeitpunkt im dysfunktionalen Bereich, wurde empfohlen, die Behandlung fortzusetzen, bei Werten im funktionalen Bereich erfolgte die Empfehlung, ein Ende der Behandlung zu erwägen. Zusätzlich erhielten auch die Patienten wöchentliche Rückmeldungen über ihren Gesundheitszustand und den Symptomverlauf. Außerdem fanden Qualitätszirkel statt, in denen die Behandler der Interventionsgruppe das Feedback und dessen Bedeutung für die weitere Therapieplanung diskutierten. In dieser Studie zeigte sich, dass auch schwer beeinträchtigte Patienten mit der computergestützten Beantwortung der

Fragebögen sehr gut zu Recht kamen und das Feedback positiv beurteilten (Puschner et al., 2009). Allerdings ließ sich kein Effekt der Feedbackintervention auf das Behandlungsergebnis oder den Symptomverlauf nachweisen.

Eine Studie, die nach Abschluss der Datenerhebung für die vorliegende Arbeit in zwei psychosomatischen Fachkliniken durchgeführt wurde, replizierte das Design der Studien der Lambert-Arbeitsgruppe in Deutschland (Probst et al. 2013, Probst et al. 2014). Es erfolgte eine Randomisierung von 436 Patienten in die Interventionsgruppe (wöchentliche Feedbacknachrichten, die ausgedruckt und im verschlossenen Umschlag an die Therapeuten übergeben wurden) oder Kontrollgruppe (kein Therapeutenfeedback). Alle Patienten beantworteten wöchentlich in Papier-Bleistift-Form den OQ-45 und das Instrument für die „Clinical Support Tools“. Zur Berechnung von Algorithmen für erwartete Verlaufskurven wurde das Programm OQ-Analyst mit den US-amerikanischen Normdaten verwendet. 184 Patienten ohne Eingangsscore oder mit weniger als drei Messzeitpunkten wurden von der Studie ausgeschlossen, da sonst die Effekte von Feedback nicht hätten untersucht werden können. Die Autoren berichten für die Subgruppe der NOT-Patienten (n=43) einen signifikanten Feedback-Effekt ( $d=0.54$ ), wobei die NOT-Patienten der Interventionsgruppe bei Therapiebeginn signifikant stärker beeinträchtigt waren als die NOT-Patienten der Kontrollgruppe. Für die Mehrheit der 209 Patienten, die als „on track“ – also ohne auffälligen Behandlungsverlauf – klassifiziert wurden, konnte kein Effekt des Therapeutenfeedbacks gefunden werden (Probst et al., 2014). Der Effekt der Feedbackintervention für die Gesamtstichprobe wird nicht berichtet. Da jedoch nur für die Subgruppe der NOT-Fälle ein signifikanter Effekt gefunden wurde, kann angenommen werden, dass es keinen signifikanten Haupteffekt gab.

### **1.5.2.2 Ambulantes Setting**

**Die „TK-Studie“.** Das Modellprojekt der Techniker Krankenkasse „Qualitätsmonitoring in der ambulanten Psychotherapie“ – „TK-Studie“ – verglich das herkömmliche Antrags- und Gutachterverfahren mit einer Intervention, die ein fragebogengestütztes Monitoring mit Therapeutenfeedback sowie ein modifiziertes Gutachterverfahren umfasste (Lutz, Wittmann, Böhnke, Rubel, & Steffanowski, 2012). Zu Therapiebeginn, bei Therapieende und zur 12-Monats-Katamnese wurden Patientendaten (BSI, IIP-D, ein störungsspezifisches Instrument) erhoben, die Therapeuten füllten bei Beginn und Ende der Therapie eine Basisdokumentation aus. In der Interventionsgruppe erfolgten zu festgelegten Zeitpunkten Verlaufsmessungen, deren Ergebnisse an die Therapeuten zurückgemeldet wurden (Kurze Zusammenfassung und Verlaufsgraph für jedes Instrument). Signale wurden nicht verwendet, auch wurden keine Handlungsempfehlungen ausgesprochen. Ein Gutachterverfahren wurde in der Interventionsgruppe nur eingeleitet, wenn zu Therapiebeginn weder eine ICD-Diagnose

gestellt wurde noch die Ergebnisse der psychometrischen Instrumente im klinisch auffälligen Bereich lagen. Ansonsten wurde für jede Behandlung das Kontingent einer Langzeittherapie zur Verfügung gestellt.

Das clusterrandomisierte Design teilte den beiden Gruppen je 200 Therapeuten zu, die die freiwillige Studienteilnahme den insgesamt 4.452 TK-Versicherten anboten, die sich im Rekrutierungszeitraum bei ihnen in psychotherapeutische Behandlung begaben. 1.708 Patienten willigten in die Studienteilnahme ein. Bei Therapieende lagen Daten von 597 Patienten vor und vollständige Datensätze (Therapiebeginn, Therapieende, Katamnese) existierten für insgesamt 300 Patienten, so dass die Drop-Out-Rate durch den wissenschaftlichen Projektbeirat als erheblich höher als in vergleichbaren Studien bewertet wurde (Scheidt et al., 2012). Aufgrund fachpolitischer Interessen waren nach Aussage der Evaluatoren Kompromisse hinsichtlich des Designs notwendig, die eine differenzierte Betrachtung der Wirkung von Rückmeldungen auf das Therapieergebnis und die Therapiedauer nicht ermöglicht hätten (Lutz, Wittmann, Böhnke, Rubel, & Steffanowski, 2013): Zum einen fehlte eine Kontrollgruppe mit Verlaufsmessungen ohne Therapeutenfeedback (Für die Kontrollgruppe liegen lediglich Daten zu Therapiebeginn, Therapieende und Katamnese vor). Zum anderen waren die Behandlungskontingente beider Gruppen unterschiedlich, womit die Studienleiter die mit durchschnittlich 42,8 Sitzungen (Median 43) signifikant längere Therapiedauer in der Interventionsgruppe gegenüber durchschnittlich 36,1 Sitzungen (Median 29) in der Kontrollgruppe erklären (Lutz et al., 2012).

Bezüglich der Ergebnisqualität (Effektivität) konnte kein Unterschied zwischen Interventions- und Vergleichsgruppe festgestellt werden (Lutz et al., 2012). Als wichtigsten Nebenbefund des Projekts wertet der wissenschaftliche Beirat, dass sich in beiden Gruppen sehr deutliche Verbesserungen sowohl in den Einzelmaßen als auch im multiplen Ergebnismaß (Effektgrößen  $d > 1,0$ ) zeigten (Scheidt et al. 2012).

**Das „QS-PSY-BAY“-Projekt.** Das Projekt „Qualitätssicherung in der ambulanten Psychotherapie in Bayern“ („QS-PSY-BAY“, Steffanowski et al. 2012) als ein weiteres großes Projekt der patientenzentrierten Psychotherapieforschung in Deutschland hatte zum Ziel, ein empirisch basiertes QS-System zu entwickeln, das mit Hilfe eines computergestützten Dokumentationssystems Prozess- und Ergebnisqualität abbildet. Patientendaten - Basisdokumentation, Beeinträchtigung (PHQ-D), Lebensqualität (SEL-K) und die therapeutische Beziehung (HAQ) - wurden elektronisch erfasst und an die Behandler zurückgemeldet. Erhebungen fanden bei Therapiebeginn, nach fünf probatorischen Sitzungen, bei Therapieende und zum 1-Jahres-Follow-up statt. Daten von 1696 Patienten, die von 217 Psychotherapeuten behandelt wurden, wurden zwischen 2007 und 2009 erhoben. Vollständige Daten lagen für 416 Patienten bei Therapieende und für 306 Patienten zum Follow-up-Zeitpunkt vor. Das Projekt erzielte eine sehr hohe Akzeptanz bei den beteiligten

Patienten und ergab bezüglich der Ergebnisqualität hohe Effektstärken – für die Intent-to-treat-Stichprobe  $d=0,97$  und für Patienten, die an allen Erhebungen teilgenommen hatten (per protocol)  $d=1,57$  bis Therapieende und  $d=1,59$  bis follow-up. Bereits im Zeitraum zwischen Therapiebeginn und Ende der probatorischen Sitzungen zeigte sich ein mittlerer Effekt von  $d=0,46$ . Da das Design keine Kontrollgruppe ohne Therapeutenfeedback vorsah, konnte in diesem Projekt der Effekt des – ausschließlich frühen - Verlaufsfeedbacks allerdings nicht untersucht werden.

### 1.5.3 Studien anderer Arbeitsgruppen

Das Design der Feedbackstudien der Lambert-Arbeitsgruppe (Verwendung des OQ-45, Übernahme des Bewertungsalgorithmus zur Bestimmung von „not on track“-Patienten, z.T. Einsatz von CST) wurde in weiteren Studien mit unterschiedlichen Behandlungssettings repliziert. Simon, Lambert, Harris, Busath und Vazquez (2012) führten in derselben Klinikambulanz wie Hawkins et al. (2004) eine Replikationsstudie durch, die einen kleinen, jedoch statistisch signifikant Effekt ergab ( $d=0.16$ ). Crits-Christoph et al. (2012) fanden einen positiven Effekt hinsichtlich reduziertem Drogen- und Alkoholkonsum für NOT-Patienten mit Substanzabhängigkeit. Ein signifikanter Effekt der Feedbackintervention bzgl. der OQ-Gesamtwert-Differenz wird außerdem für essgestörte Patientinnen in stationärer Behandlung berichtet (Simon et al., 2013). Auch eine Studie im Kontext ambulanter Paartherapie zeigte signifikant häufiger Verbesserungen für die Feedback-Bedingung, die bei follow-up stabil blieben (Anker, Duncan, & Sparks, 2009). In einer norwegischen Studie, die sowohl ambulante als auch stationäre Patienten einschloss, zeigte sich ebenfalls ein signifikanter Effekt der Feedbackintervention für die Gesamtstichprobe ( $d=0.32$ , Amble, Gude, Stubdal, Anderson, & Wampold, 2015). Für stationäre Therapiepatienten, die ausschließlich im Gruppensetting behandelt worden waren, zeigte sich ein positiver Effekt von Patienten- und Therapeutenfeedback bzgl. depressiver Symptomatik, obwohl nur zu einem Zeitpunkt im Therapieverlauf Feedback gegeben worden war (Newnham, Hooke, & Page, 2010). Allerdings war dieser Effekt nur auf drei Subskalen zu beobachten, bezüglich fünf weiterer Skalen sowie des Instruments, das zur Verlaufsmessung eingesetzt worden war (WHO-5), zeigten sich keine Unterschiede.

Der Effekt von Verlaufsfeedback ist allerdings nicht so robust, wie es die Mehrzahl der dargestellten Befunde nahelegen. In einer Studie an ambulanten Psychotherapiepatienten aus den Niederlanden ( $n=475$ ), die das Design der Lambert-Gruppe replizierte, zeigten sich weder für Therapeutenfeedback noch für kombiniertes Patienten- und Therapeutenfeedback signifikante Unterschiede bzgl. des Therapieergebnisses im Vergleich zu TAU (de Jong et al., 2014). Eine Meta-Analyse von Knaup, Kösters, Schöfer, Becker und Puschner (2009) schloss neben vier Studien aus der Lambert-Gruppe noch acht weitere Studien ein (zwei aus

Deutschland - Bauer, 2004; Berking et al., 2006, vgl. 1.5.2.1). Bezüglich der kurzfristigen Effektivität (Therapiebeginn bis Therapieende) ergab sich eine Effektgröße von  $d=0.10$  (95% KI: 0.01-0.19). Damit zeigte sich ein kleiner, jedoch statistisch signifikanter Effekt von Verlaufsfeedback auf das kurzfristige Behandlungsergebnis. Bei Betrachtung des längerfristigen Outcomes (3-12 Monate nach initialer Messung), das lediglich in 5 Studien untersucht worden war, zeigte sich allerdings kein Effekt von Verlaufsfeedback. Ein aktueller systematischer Review, der allerdings einige der berichteten Studien ausgeschlossen hatte (z.B. wenn andere Outcome-Maße als der OQ verwendet worden waren), kommt bei Betrachtung der Gesamtstichproben sogar zu dem Schluss, dass sich zwischen Gruppen mit und ohne Verlaufsfeedback keine Unterschiede zeigen (Kendrick et al., 2016).

#### 1.5.4 Differenzielle Effekte von Verlaufsfeedback in der Psychotherapie

##### **1.5.4.1 Auffällige Verläufe („not on track“-Fälle)**

Es gibt nur wenige Studien, die einen vergleichbaren Effekt von Therapeutenfeedback sowohl bei „on track“ als auch bei „not on track“-Patienten zeigen (Bickman, Douglas Kelley, Breda, de Andrade, & Riemer, 2011; Hawkins et al., 2004; Simon et al., 2013; Amble et al., 2014). Feedback hatte in vielen Studien keinen Effekt bei „on track“-Patienten, die den erwarteten Therapiefortschritt zeigen (Crits-Christoph et al., 2012; Harmon et al., 2007; de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; Lambert, Whipple et al., 2001; Lambert, Whipple, Vermeersch et al., 2002; Newnham et al., 2010; Probst et al., 2013; Simon et al., 2012; Slade et al., 2008; Whipple et al., 2003). Meta-Analysen und Reviews zeigen, dass Feedback vor allem bei Patienten, deren Symptomatik sich anfänglich verschlechtert oder nicht verändert („not on track“), zu einem besseren Behandlungsergebnis führt (Carlier et al. 2012, Knaup et al., 2009; Lambert & Shimokawa, 2011, Shimokawa et al., 2010).

Interessanterweise entfaltet zwar Feedback bei NOT-Patienten seine stärkste (und oft auch ausschließlich dort) Wirkung, scheint aber häufig nicht geeignet zu sein, NOT-Verläufe zu reduzieren (Amble et al., 2015). Dieser Befund konnte in post-hoc Analysen von Daten aus 10 Studien, die sich auf das Design der Lambert-Arbeitsgruppe stützen, repliziert werden: In keiner Studie gab es signifikante Unterschiede bezüglich des Anteils an NOT-Patienten in Gruppen mit und ohne Verlaufsfeedback (Amble et al., 2015). Damit übereinstimmend berichten Probst et al. (2013) - die ebenfalls einen signifikanten Feedbackeffekt für NOT-Patienten fanden - dass sich auch in der Feedbackbedingung NOT-Patienten bis zum Behandlungsende verschlechtern, nur in geringerem Ausmaß als in der Kontrollbedingung. Bei de Jong et al. (2014) konnte in Kurzzeittherapien bei NOT-Fällen mit Verlaufsfeedback die (weitere) negative Veränderung, die bei NOT-Fällen ohne Feedback auftrat, aufgehalten werden – allerdings ohne eine Verbesserung der Symptomatik zu erreichen.

#### **1.5.4.2 Differenzielle Allokation der Behandlungsdauer**

Die Behandlungsdauern von OT- und NOT-Patienten unterschieden sich in den Studien der Lambert-Arbeitsgruppe in Abhängigkeit davon, ob die Behandler Verlaufsfeedback erhalten hatten oder nicht. Bei Lambert, Whipple et al. (2001) erhielten NOT-Patienten in der Feedback-Bedingung durchschnittlich 9,7 Therapiesitzungen, in der Bedingung ohne Feedback dagegen nur 5,0 Sitzungen. Demgegenüber wurden OT-Patienten in der Feedback-Bedingung etwas kürzer behandelt (M=2,8 Sitzungen) als in der Bedingung ohne Feedback (M=3,6 Sitzungen), so dass die Gesamtbehandlungsdauer in den Gruppen mit und ohne Feedback mit durchschnittlich 3,6 bzw. 3,7 Sitzungen vergleichbar war. „...the main effect of feedback was to keep the NOT-feedback patients in treatment longer“, lautet die Erklärung des Wirkmechanismus durch die Autoren (Lambert, Whipple et al., 2001, p. 63). Das Ergebnis einer Effizienzsteigerung (längere Behandlungsdauer für NOT-Patienten, kürzere für OT-Patienten) konnte in darauffolgenden Studien repliziert werden. In der Meta-Analyse von Shimokawa et al. (2010) wird bezüglich der durchschnittlichen Sitzungszahl der NOT-Patienten eine aggregierte Effektstärke von  $g=0.27$  zugunsten der Feedbackbedingungen berichtet. Dies entspricht von der Größe her der statistisch signifikanten Effektstärke bezüglich des Therapieergebnisses, erreicht allerdings aufgrund der großen Heterogenität zwischen den Studien nicht statistische Signifikanz. In der Studie, die die größte Effektstärke hinsichtlich des Behandlungsergebnisses aufwies, zeigte sich auch der größte Unterschied in der Behandlungsdauer zwischen NOT-Patienten mit und ohne Therapeutenfeedback (Lambert, Whipple et al., 2001). Eine längere Behandlungsdauer von NOT-Patienten, deren Therapeuten Verlaufsfeedback erhielten im Vergleich zu NOT-Patienten ohne Therapeutenfeedback berichten auch Probst et al. (2013). Des Weiteren zeigt eine Auswertung *aller* NOT-Patienten (mit und ohne Feedbackinterventionen), dass diejenigen, die als „early terminators“ klassifiziert wurden (Behandlungsende nach 5 oder weniger Sitzungen) im Vergleich zu „late terminators“ (sechs und mehr Sitzungen) ein 2.24 höheres Risiko für reliable Verschlechterungen aufwiesen. „These findings underline the need to retain NOT patients in treatment longer“, schlussfolgern die Autoren aus diesen Ergebnissen (Shimokawa et al., 2010, p.309).

Ein Wirkmechanismus von Verlaufsfeedback scheint demnach zu sein, dass Patienten, deren Symptomverlauf keine Veränderungen oder Verschlechterungen aufweist, länger in Behandlung „gehalten“ werden als vergleichbare NOT-Patienten, deren Behandler kein Verlaufsfeedback erhalten. Obwohl das bei Lambert und Kollegen verwendete Feedbacksystem auf dem Dosis-Wirkungs-Modell basiert und bei ausbleibender Response ein Alarmsignal sendet, verbunden mit verschiedenen Empfehlungen (wobei eine Verlängerung der Behandlung bei schlecht respondierenden Patienten explizit nicht dazu gehört!), scheinen sich die Behandler implizit so zu verhalten, wie es die Feedbackstrategie nach dem Random Walk Modell empfiehlt (vgl. 4.2.1). Bauer (2004) konnte eine höhere Effizienz (kürzere



Behandlungsdauer bei vergleichbarem Ergebnis) durch Verlaufsfeedback zeigen, allerdings nur unter der Bedingung, dass bei Symptomwerten im „gesunden“ Normbereich explizit die Empfehlung gegeben wurde, die Behandlung zu beenden.

Trotz fehlender Kontrollgruppe mit Verlaufsmessungen sehen die an der TK-Studie beteiligten Wissenschaftler ebenfalls Hinweise auf eine stärkere adaptive Allokation der Behandlungsdauer im TK-Modell: Patienten, deren Symptomatik sich bis zur ersten Zwischenmessung (10. Therapiesitzung) verschlechterte, erhielten mehr Sitzungen als Patienten mit einer initialen Symptomreduktion, auch wenn für die Anzahl an bewilligten Sitzungen kontrolliert wurde (Lutz et al., 2012).

Allerdings gibt es auch Studien, bei denen sich die Behandlungsdauern von OT- und NOT-Patienten in den Feedbackbedingungen im Vergleich zu TAU nicht unterschieden (Hawkins et al., 2004; Crits-Christoph et al., 2012), was für weitere Wirkmechanismen von Verlaufsfeedback spricht. Bei Newnham et al. (2010), die einen kleinen positiven Feedbackeffekt berichten, war die Behandlungsdauer aufgrund des Settings (zwei Wochen stationäre Gruppentherapie) für alle Patienten konstant. Weiterhin gibt es Studien, die die Behandlungsdauer nicht für OT und NOT-Fälle getrennt aufschlüsseln, sondern nur die Behandlungsdauer für die Gesamtgruppen berichten (de Jong et al., 2014; Puschner et al., 2009; Amble et al., 2015; Simon et al., 2013)

#### **1.5.4.3 Feedbackergänzungen: Patientenfeedback und „Clinical Support Tools“**

Die Befunde sprechen überwiegend dafür, dass ein kombiniertes Patienten- und Therapeutenfeedback dem ausschließlichen Feedback an Therapeuten nicht überlegen ist. Knaup et al. (2009) berichten zwar eine etwas höhere Effektivität für die Kombination von Patienten- und Therapeutenfeedback, Shimokawa et al. (2010) konnten jedoch keinen zusätzlichen Effekt feststellen, wenn neben den Behandlern auch die Patienten Feedback erhielten. Auch bei de Jong et al. (2014) zeigten sich in der Gesamtstichprobe keine Unterschiede zwischen ausschließlichem Therapeutenfeedback und kombiniertem Patienten- und Therapeutenfeedback. Dagegen erwies sich der (kleine) zusätzliche Effekt, wenn Feedback bei NOT-Fällen um die „clinical support tools“ ergänzt wurde, bei Shimokawa et al. (2010) als stabil ( $g=-0.16$ ). Allerdings wurden NOT-Patienten in der CST-Feedbackintervention signifikant länger behandelt als NOT-Patienten in der „normalen“ Feedbackintervention, so dass der zusätzliche Effekt bezüglich des Therapieergebnisses durch die längere Therapiedauer „konfundiert“ sein könnte (vgl. 1.5.4.2). Damit konform zeigten sich keine Unterschiede in den Behandlungsdauern von NOT-Patienten mit ausschließlich Therapeutenfeedback und kombiniertem Therapeuten- und Patientenfeedback (Shimokawa et al., 2010),

#### **1.5.4.4 Behandlungssetting und Patientenmerkmale**

**Kurzzeit -vs. Langzeittherapien.** Studien, die einen positiven Effekt von Verlaufsfeedback auf das Behandlungsergebnis berichten, beziehen sich meist auf therapeutische Settings mit kurzer Behandlungsdauer. Gleichzeitig gibt es Befunde, dass Verlaufsfeedback auch im stationären Setting wirksam ist (Berking et al., 2006; Newnham, et al., 2010), bei Newnham et al. (2010) war die Behandlungsdauer mit a priori festgelegten zwei Wochen allerdings ebenfalls kurz.

In Langzeittherapien könnten andere Wirkmechanismen bedeutsam sein als in Kurzzeittherapien. Bei de Jong et al. (2014) zeigten sich Unterschiede zwischen OT-Fällen und NOT-Fällen in Kurzzeittherapien, jedoch nicht in Langzeittherapien. Gleichzeitig war ausschließlich in Langzeittherapien ein kombiniertes Patienten- und Therapeutenfeedback den Bedingungen „kein Feedback“ und „nur Therapeutenfeedback“ leicht überlegen (allerdings nur bezüglich der Veränderungsgeschwindigkeit, nicht bezüglich des Therapieergebnisses). In Langzeittherapien haben Patienten eventuell bereits genügend Kontingenz zur Verfügung, um eine ausbleibende „early response“ zu kompensieren und „aufzuholen“. Hier könnten aber andere Mechanismen von Verlaufsfeedback greifen (z.B. Commitment des Patienten stärken oder Therapiemotivation aufrechterhalten).

**Beeinträchtigungsschwere.** Bezüglich des moderierenden Einflusses der (initialen) Beeinträchtigungsschwere erscheinen die Befunde widersprüchlich. Aufgrund der positiven Effekte der „frühen“ Studien der Lambert-Arbeitsgruppe, die alle ein Studentenklientel mit guter Alltagsfunktionalität, geringer Symptomschwere und kurzer Krankheitsdauer untersuchten, war angenommen worden, dass Feedback effektiver sei, wenn das Patientenklientel weniger stark beeinträchtigt ist. Das schlussfolgerten z.B. Simon et al. (2012), die stärker beeinträchtigte Patienten einer Klinikambulanz untersuchten und einen deutlich kleineren Feedbackeffekt fanden als die Studien, die in der Studierendenberatung durchgeführt worden waren. Auf der anderen Seite wurden in Studien, in denen sich keine differenziellen Effekte für OT und NOT-Fälle zeigten, Patienten mit hohen initialen Belastungswerten untersucht und gute Feedbackeffekte erzielt, was die Autoren schlussfolgern ließ, dass Feedback bei initial stärker beeinträchtigten Patienten wirksamer sein könnte (Hawkins et al., 2004; Simon et al., 2013; Amble et al., 2015).

#### **1.5.4.5 Therapeutenvariablen**

Bisher wurde die Rolle der Behandler, die meist die einzigen Adressaten der Rückmeldungen sind, kaum untersucht. Es gibt jedoch Hinweise, dass Therapeutenvariablen die Effekte von Verlaufsfeedback moderieren. So scheint Feedback nur bei einem Teil der Therapeuten zu einer Verbesserung der Therapieergebnisse zu führen (Simon et al., 2012). De Jong et al. (2012) fanden weder einen signifikanten Feedbackeffekt für die Gesamtgruppe noch für die

Subgruppe der NOT-Patienten, allerdings zeigte Verlaufsfeedback dann einen positiven Effekt bei NOT-Fällen, wenn Therapeuten angaben, das Feedback aktiv zu nutzen (was bei 46% der Therapeuten der Fall war). Eine aktive Feedbacknutzung wiederum war häufiger bei weiblichen Therapeuten und Behandlern, die eine größere Bereitschaft zur Feedbacknutzung mitbrachten („commitment“). Patienten von Therapeuten, die Verlaufsfeedback erhielten, zeigten außerdem ein größeres Ausmaß an Veränderung, wenn ihre Therapeuten eine höhere Selbstwirksamkeitserwartung hatten (De Jong et al., 2012).

Riemer und Bickman (2011) berichten, dass hoher administrativer Arbeitsaufwand als wichtigster Hinderungsgrund, Feedback aktiv zu nutzen, genannt wird, so dass Therapeutenvariablen (wie Feedbacknutzung) wiederum durch Struktur-/Organisationsvariablen beeinflusst werden können. Außerdem scheinen sich Therapeutenunterschiede bezüglich der Behandlungsergebnisse mit zunehmender Beeinträchtigungsschwere der Patienten zu verstärken (Saxon & Barkham, 2012), so dass moderierende Effekte von Patienten- und Therapeutenmerkmalen auf Verlaufsfeedback miteinander interagieren könnten.

## 1.6 Die therapeutische Beziehung

Als die Effektivität verschiedener psychotherapeutischer Verfahren durch zahlreiche Studien nachgewiesen worden war, gewann die Konzeptualisierung und Untersuchung von Schulen übergreifenden Wirkfaktoren, den sog. „common factors“, an Bedeutung. Damit werden Elemente bezeichnet, die in allen therapeutischen Schulen und Orientierungen eine Rolle spielen, jedoch nicht notwendigerweise als zentrale Faktoren dieser Ansätze beschrieben werden. Es wird davon ausgegangen, dass die „common factors“ mit Schätzungen bis zu 50% den Löwenanteil der Varianz bzgl. des Therapieergebnisses erklären – deutlich mehr als therapiespezifische Methoden mit ca. 10% bis 15% (Cuijpers et al., 2012; Lambert, 2013). Wampold und Imel (2015) benennen die therapeutische Beziehung - neben den Erfolgserwartungen des Patienten und einer empathischen Verbindung zwischen Patient und Therapeut – als bedeutenden „allgemeinen“ Wirkfaktor in der Psychotherapie.

### 1.6.1 Das Konzept der therapeutischen Beziehung

Die Bedeutung der Allianz zwischen Patient und Therapeut wurde bereits von Freud (1912, 1913) betont. In der Psychoanalyse ist die therapeutische Allianz eng mit dem Konzept der Übertragung verbunden, wobei bereits Freud die therapeutische Allianz von der Übertragung abgrenzt. Weitere Meilensteine für die Entwicklung des Konzepts der therapeutischen Allianz in der Psychoanalyse legten z.B. Sterba (1934), Zetzel (1956) und Greenson (1965), bis schließlich versucht wurde, die therapeutische Beziehung zu operationalisieren und empirisch zu erfassen (z.B. Luborsky 1976). Thomä und Kächele (2006) unterscheiden als Komponenten

der Beziehung in der Psychoanalyse die Aspekte Übertragung, Realbeziehung und therapeutische Allianz.

Auch wenn verschiedene Therapieschulen andere Annahmen für eine erfolgreiche Therapie voraussetzen als die Psychoanalyse, wird der therapeutischen Beziehung auch dort eine wichtige Bedeutung beigemessen. Dabei stellt sich grundsätzlich die Frage, ob die therapeutische Beziehung lediglich als Basiskompetenz angesehen wird, die den Rahmen für die Anwendung therapeutischer Techniken bildet, oder als eine therapeutische Grundhaltung im Sinne eines psychotherapeutischen Wirkfaktors (wie dies in der Gesprächspsychotherapie der Fall ist) oder aber als eine therapeutische Technik selbst (Lammers, 2017). Nachdem die therapeutische Beziehung in der Verhaltenstherapie lange Zeit vernachlässigt worden war, gewann sie mit der Konzeption der „vertikalen Verhaltensanalyse“ (Grawe, 1982) an Bedeutung. Eine erfolgreiche Therapiebeziehung soll durch komplementäre Beziehungsgestaltung hergestellt werden (Grawe, 1992; Znoj, 2005), wobei auf den Bedürfnisaspekt problematischer Verhaltensweisen fokussiert wird anstatt auf das Problemverhalten selbst zu reagieren. In sog. „Dritte Welle Verfahren“ der Verhaltenstherapie kommt der therapeutischen Beziehung eine wichtige Rolle im Rahmen der therapeutischen Methoden zu (z.B. „Diszipliniertes Persönliches Einlassen“ im Cognitive Behavioral Analysis System of Psychotherapy CBASP, McCullough, 2006 oder „begrenztes Nachbeeltern“ in der Schematherapie, Young, Klosko, & Weishaar 2003).

Es überrascht nicht, dass die Beziehung zwischen Patient und Therapeut in den Fokus der „Common Factors“-Forschung rückte. Bordin (1979) schlug das ursprünglich psychoanalytische Konzept der therapeutischen Allianz als ein Wirkprinzip vor, das allen psychotherapeutischen Verfahren gemeinsam ist. Seine Definition von therapeutischer Allianz enthält die drei Komponenten Übereinstimmung von Patient und Therapeut bezüglich der Therapieziele („goals“), therapeutische Aufgaben („tasks“) und die Entwicklung einer Bindung zwischen Patient und Therapeut („bond“). Damit wurde das in der Psychoanalyse auf die Bindung fokussierte Konzept um den Aspekt der Arbeitsbeziehung erweitert. Da sowohl Ziele als auch Aufgaben je nach Therapieverfahren spezifisch formuliert werden können, ist mit Bordins Formulierung eine theorieunabhängige Definition gegeben.

### 1.6.2 Die Bedeutung der therapeutischen Beziehung für das Therapieergebnis

In vielen Studien zeigte sich eine signifikante Beziehung zwischen der therapeutischen Beziehung und dem Therapieergebnis (Orlinsky, Ronnestad, & Willutzki, 2004). Horvath und Symonds (1991) berichten in ihrer Meta-Analyse (24 Studien) Korrelationsschätzungen zwischen der therapeutischen Allianz und dem Therapieerfolg von  $r=0,29$ . Horvath und Bedi (2002) geben Schätzungen zwischen 0,21 und 0,29 an. Martin, Garske und Davis (2000) berichten in ihrer Meta-Analyse (79 Studien) eine durchschnittliche Korrelation zwischen

therapeutischer Beziehung und dem Therapieergebnis von  $r=0,22$ . Die mit  $r=0,64$  höchste Korrelation zwischen therapeutischer Beziehung und dem Behandlungsergebnis (beides erfasst aus Therapeutensicht) in einer deutschsprachigen Studie berichten Rudolf und Manz (1993). In einer Studie des National Institute of Mental Health zur Effektivität unterschiedlicher Therapien bei der Behandlung von Depression (Krupnick et al., 1996) beeinflusste die therapeutische Beziehung den Therapieerfolg nicht nur in den Psychotherapie-Bedingungen, sondern auch in den Bedingungen ohne Psychotherapie.

Betrachtet man die Ergebnisse der Meta-Analysen, sind die Zusammenhänge zwischen der therapeutischen Beziehung und dem Behandlungsergebnis als niedrig bis allenfalls moderat einzustufen. Dennoch gilt die therapeutische Beziehung nach wie vor als das Prozessmaß, das als Einzelvariable den größten Zusammenhang mit dem Therapieergebnis aufweist (Hentschel 2005, Lambert & Barley, 2001).

Eine Schwäche vieler Studien, die den Einfluss der therapeutischen Beziehung auf das Therapieergebnis untersuchen, ist eine Erfassung der therapeutischen Beziehung im Behandlungsverlauf, ohne eine vorherige Symptomveränderung, die sowohl die therapeutische Beziehung als auch das Behandlungsergebnis beeinflussen könnte, zu berücksichtigen. Daher ist es notwendig, die Veränderung von therapeutischer Allianz und Symptomatik im zeitlichen Verlauf sowie deren Interdependenz zu untersuchen. In einigen Studien wurde gezeigt, dass die Einschätzung der therapeutischen Beziehung positiv mit einer vorhergehenden Symptomverbesserung korrelierte (Gaston, Marmar, Gallagher & Thompson, 1991; Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000; DeRubeis & Feeley, 1990; Puschner, Wolf, & Kraft, 2009). Wurden das Ausmaß der initialen psychischen Beeinträchtigung und eine frühe Symptomveränderung kontrolliert, zeigte sich in den Studien von Gaston et al. (1991) und Barber, Crits-Christoph und Luborsky (1996) kein Zusammenhang zwischen der therapeutischen Beziehung und dem Therapieergebnis. Einige Studien zeigten, dass die therapeutische Beziehung kein Prädiktor für nachfolgende Symptomveränderung war, wenn die dem Messzeitpunkt vorhergehende Symptomveränderung kontrolliert wurde (DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999; Puschner et al., 2009). Diese Ergebnisse führten dazu, dass die Bedeutung der therapeutischen Beziehung für den Therapieerfolg infrage gestellt und „alliance as an artefact“ (Horvath & Luborsky, 1993) angesehen wurde. Im Gegensatz zu den genannten Studien war bei Barber et al. (2000) und bei Klein et al. (2003) die therapeutische Beziehung ein signifikanter Prädiktor für nachfolgende Symptomveränderung auch dann, wenn die dem Messzeitpunkt vorhergehende Symptomveränderung kontrolliert wurde. Die Befunde sprechen demnach dafür, dass die therapeutische Beziehung sowohl ein Prädiktor für als auch eine Konsequenz von Symptomveränderungen im Psychotherapieverlauf sein kann.

### 1.6.3 Die psychometrische Erfassung der therapeutischen Beziehung

Einen der ersten Versuche, die therapeutische Beziehung empirisch zu erfassen, unternahm Luborsky mit der Entwicklung der Penn Helping Alliance Scales (PennHAS), einer Rating-Methode durch externe Beobachter (Alexander & Luborsky, 1986). Dabei bilden die Wahrnehmung des Therapeuten als hilfreich und unterstützend (*Beziehungsaspekt*) sowie die Wahrnehmung, auf ein gemeinsames Ziel hinzuarbeiten (*Aspekt der Arbeitsbeziehung* nach Bordin) die beiden zentralen Aspekte der Allianz. Da Rating-Methoden sehr aufwändig sind – sie erfordern Ton- oder Videoaufzeichnungen der Therapiesitzungen – sind Fragebögen zur Erfassung der therapeutischen Beziehung deutlich weiter verbreitet. Zu den international am häufigsten eingesetzten Instrumenten zählen der Penn Helping Alliance Questionnaire (HAQ, Luborsky, 1984, deutsche Version von Bassler et al., 1995) mit den beiden oben genannten Faktoren, das Working Alliance Inventory (WAI, Horvath & Greenberg, 1989), das auf Bordin's Konzeptualisierung zurückgeht sowie die California Psychotherapy Alliance Scale (CALPAS, Gaston & Marmar, 1993). Die genannten Instrumente sind sowohl aus Patienten- als auch aus Therapeutenperspektive zu bearbeiten; ebenso liegen Versionen für externe Beobachter vor. Bezüglich der Vergleichbarkeit der verschiedenen Fragebögen berichten die meisten Studien einen gemeinsamen Varianzanteil von maximal 50 bis 60 Prozent, so dass die Skalen nicht als äquivalent anzusehen sind (Hentschel, 2005). Korrelationen zwischen Patienten- und Therapeutenratings der therapeutischen Beziehung sind mäßig (Cecero, Fenton, Frankforter, Nich, & Carroll, 2001). Daher könnte die Rückmeldung der therapeutischen Beziehung aus Patientensicht im Therapieverlauf für die Behandler hilfreich sein, um eine „schlechte“ Beziehung oder Verschlechterungen im Verlauf frühzeitig (oder überhaupt) zu erkennen. Brüche in der therapeutischen Beziehung – die sog. „alliance ruptures“ – sind im Therapieverlauf mit 19%- 42% relativ häufig (Safran, Muran, Samstag, & Stevens, 2000), so dass Verlaufsfeedback zur therapeutischen Beziehung Behandler dabei unterstützen könnte, das eigene Verhalten besser auf die Bedürfnisse des Patienten abzustimmen.

### 1.7 Verlaufsfeedback von Prozessmaßen in der Psychotherapie

In der bisher einzigen Studie, die auf Feedback bezüglich des therapeutischen Prozesses fokussierte (mit den Dimensionen Erfolgserwartungen, Empathie und therapeutische Beziehung, McClintock et al., 2017) – allerdings ohne Vergleichsgruppe mit symptombezogenem Feedback -, konnten in der Feedbackbedingung im Vergleich zu TAU stärkere Verbesserungen bezüglich der von den Patienten wahrgenommenen Empathie der Behandler und der therapeutischen Beziehung erzielt werden. Dagegen zeigten sich keine Unterschiede bezüglich der Erfolgserwartung (Outcome Expectation Questionnaire OEQ) und der Ergebnismaße (Beck Depression Inventory-II BDI-II, Schwartz Outcome Scale-10 SOS-10). Die Autoren vermuten aufgrund ihres Ergebnisses, dass Feedback ausschließlich oder

vorrangig auf die Parameter wirkt, die es auch enthält und zurückmeldet: „outcome feedback may primarily affect outcome, and process feedback may primarily affect process“ (McClintock et al., 2017, p. 256).

## 1.8 Einordnung der Arbeit in den Forschungsstand

Zur Effektivität von Verlaufsfeedback in der Psychotherapie liegen mittlerweile zahlreiche Studien und auch Meta-Analysen vor. Allerdings stammen die meisten Studien nach wie vor aus einer einzigen Arbeitsgruppe (Lambert und Kollegen) mit sehr spezifischen, auf das hiesige System nicht übertragbaren Bedingungen (universitäre Beratungsstelle mit eher gering belasteten Patienten, hohen Abbruchraten und ausgesprochen kurzer Behandlungsdauer) oder replizieren das Design dieser Studien. Während in den „frühen“ Studien durchweg positive Effekte von Verlaufsfeedback auf das Therapieergebnis berichtet wurden, wird die empirische Evidenz aktuell zurückhaltender beurteilt. Lambert (2015) betrachtet ausschließlich die Effektstärke für die Minderheit der „not on track“-Fälle und bewertet diese mit durchschnittlich 0.5 als beträchtlich, da zwei „aktive“ Behandlungsformen verglichen werden und nicht – wie häufig in Therapiestudien – eine Therapieform mit einer Warteliste. Dagegen ziehen Lutz et al. (2012) und Kendrick et al. (2016) bei Betrachtung der Gesamtstichproben das Fazit, dass ein Feedbackeffekt entweder minimal oder nicht vorhanden ist.

Für das deutsche Versorgungssystem liegen bislang vier Studien vor. Die Arbeit von Bauer (2004) verwendete ein prospektives Kontrollgruppendesign mit sequenziellen Untersuchungsphasen; es handelte sich also nicht um eine kontrollierte randomisierte Studie. Ferner wurde nur Feedback zu einem Zeitpunkt im Therapieverlauf gegeben, das aufgrund der Papier-Bleistift-Erhebung mit Zeitverzögerung an die Therapeuten zurück gemeldet wurde und keinen Effekt hinsichtlich der Effektivität, jedoch einen kleinen Effekt bezüglich der Effizienz erzielte. Auch die zweite in Deutschland durchgeführte Studie zur Effektivität von therapiebegleitendem Feedback griff auf Papier-Bleistift-Erhebungen zurück. In dieser Studie (Berking et al., 2006) zeigte sich zwar ein Effekt der Feedbackbedingung, allerdings fehlte die Möglichkeit der Implementierung in die Routineversorgung bzw. war dies auch nicht Ziel dieser Untersuchung. Während die Studie von Puschner et al. (2009) trotz einer elaborierten Intervention (Patienten- und Therapeutenfeedback, Qualitätszirkel) keinen Effekt von Verlaufsfeedback erbrachte, zeigte sich bei Probst et al. (2014), die das Design der Lambert-Arbeitsgruppe replizierten, ein Feedbackeffekt ausschließlich bei „not on track“-Fällen. Da Letztere keine Ergebnisse für die Gesamtstichprobe berichten, ist die Vergleichbarkeit mit den anderen Studien – speziell jenen, die ausschließlich Effekte für die Gesamtstichprobe berichten (z.B. (Puschner et al., 2009) - allerdings eingeschränkt. Die empirischen Befunde bezüglich der Effektivität von Verlaufsfeedback sind demnach heterogen – speziell für das

hiesige Versorgungssystem. Daher besteht weiterhin Bedarf, die empirische Basis bezüglich der Effektivität von Verlaufsfeedback in der Psychotherapie zu erweitern.

Weiterhin steht eine systematische Überprüfung unterschiedlicher Feedbackinhalte – insbesondere Inhalte, die sich nicht ausschließlich auf die Symptomatik des Patienten beziehen – noch aus. Lambert und Kollegen ergänzten mit den „Clinical Support Tools“ zusätzliche Maße (therapeutische Beziehung, Änderungsmotivation, soziale Unterstützung) – allerdings ausschließlich bei NOT-Patienten, die nicht die erwartete Besserung der Symptomatik zeigen oder deren Symptome sich im Therapieverlauf verschlechtern. Da davon auszugehen ist, dass verschiedene Prozessvariablen den Verlauf der psychischen Symptomatik beeinflussen, erscheint es sinnvoll, Prozessmaße zu erfassen und an die Behandler zurück zu melden, bevor diese ggf. negativen Einfluss auf den Symptomverlauf ausüben. Bisher gibt es lediglich eine Studie, die die systematische Rückmeldung von Prozessaspekten auf das Therapieergebnis untersucht hat (McClintock et al., 2017). Allerdings fehlte in dieser Untersuchung eine Vergleichsgruppe mit symptombezogenem Feedback. Diese Lücke will die vorliegende Studie schließen, indem in verschiedenen Untersuchungsgruppen von Anfang an sowohl symptombezogene als auch zusätzlich prozessbezogene Verlaufsrückmeldungen gegeben werden, so dass der mögliche Nutzen einer zusätzlichen Rückmeldung der wichtigsten Prozessvariablen – der therapeutischen Beziehung – systematisch untersucht werden kann.

## 1.9 Fragestellungen und Hypothesen

Diese Arbeit untersucht die Praktikabilität, die Akzeptanz und die Effektivität von computergestützten Rückmeldungen an Bezugstherapeuten im Verlauf stationärer Psychotherapie unter Routinebedingungen. Das Feedback bezieht sich entweder ausschließlich auf die Symptombelastung und den Symptomverlauf (symptombezogenes Feedback) oder zusätzlich auf die therapeutische Beziehung (prozessbezogenes Feedback), beides erhoben aus Patientensicht.

Es werden die folgenden drei Untersuchungsgruppen verglichen:

- Patienten von Therapeuten, die ausschließlich symptombezogenes Feedback sechs Wochen nach Therapiebeginn erhalten (**Gruppe 1**)
- Patienten von Therapeuten, die ausschließlich symptombezogenes Feedback zu drei Zeitpunkten im Therapieverlauf erhalten (unmittelbar nach Aufnahme, zwei Wochen nach Aufnahme, sechs Wochen nach Aufnahme, **Gruppe 2**)
- Patienten von Therapeuten, die sowohl symptom- als auch prozessbezogenes Feedback zu drei Zeitpunkten im Therapieverlauf erhalten (unmittelbar nach Aufnahme, zwei Wochen nach Aufnahme, sechs Wochen nach Aufnahme, **Gruppe 3**)



Es werden folgende Fragestellungen und Hypothesen formuliert:

### 1.9.1 Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen von computergestütztem Therapeutenfeedback

1.9.1. a) Sind computergestützte Verlaufserhebungen und Therapeutenfeedback bei einem unselektierten Patientenkollektiv unter Routinebedingungen durchführbar?

1.9.1. b) Akzeptanz: Inwieweit stimmen die Therapeuten dem Feedback zu? Werden die computergestützten Rückmeldungen als Bestätigung, Ergänzung oder Veränderung des klinischen Eindrucks angesehen?

1.9.1. c) Relevanz und Nutzen: Wie relevant werden die Informationen der Feedbacknachrichten bewertet? Inwieweit werden die Rückmeldungen als hilfreich für die weitere Behandlung gesehen?

### 1.9.2 Fragestellung 2: Gruppenunterschiede bezüglich der Effektivität

Bei Patienten, die zwei Wochen nach Aufnahme als „auffällig“ klassifiziert werden und deren Therapeuten Feedback zu drei Zeitpunkten im Therapieverlauf erhalten (Gruppen 2 und 3), werden im Vergleich zu „auffälligen“ Patienten, deren Therapeuten ausschließlich Feedback sechs Wochen nach Aufnahme erhalten (Gruppe 1), bessere Therapieergebnisse erwartet.

#### Hypothesen:

Patienten mit einem frühen auffälligen Verlauf der Gruppen 2 und 3 zeigen gegenüber früh auffälligen Patienten der Gruppe 1 bei Behandlungsende:

1.9.2. a) Eine niedrigere globale psychische Symptombelastung (Globaler Schwereindex GSI der SCL-90)

1.9.2. b) Einen größeren Prozentsatz an reliabler Verbesserung der psychischen Symptomatik (GSI)

1.9.2. c) Einen geringeren Prozentsatz an auffälligen Verläufen

### 1.9.3 Fragestellung 3: Gruppenunterschiede in der Behandlungsdauer

Ausgehend von den unter 1.5.4 berichteten Befunden wird angenommen, dass das Feedback in sowohl früher als auch späterer Phase des Therapieverlaufs (Gruppe 2 und 3) zu einer stärkeren adaptiven Allokation der Behandlungsdauer führt, also die Behandlungsdauer stärker an der Beeinträchtigungsschwere des Patienten orientiert wird.

#### Hypothesen:

1.9.3. a) Patienten, die zwei Wochen nach Aufnahme als „auffällig“ klassifiziert werden, werden in den Gruppen 2 und 3 länger behandelt als „früh auffällige“ Patienten in Gruppe 1.

1.9.3. b) Patienten, deren psychische Gesamtbelastung (Globaler Schwereindex GSI) zwei Wochen nach Aufnahme innerhalb des Normbereichs („funktionaler“ Bereich) liegt, werden in den Gruppen 2 und 3 kürzer behandelt als Patienten in Gruppe 1.

#### 1.9.4 Fragestellung 4: Gruppenunterschiede bezüglich der therapeutischen Beziehung

Es wird ein spezifischer Effekt postuliert, d.h. ein höheres Ausmaß an Verbesserung der therapeutischen Beziehung, wenn diese im Therapieverlauf an die Therapeuten zurückgemeldet wird. Aufgrund des zu erwartenden Deckeneffekts in den Ratings der therapeutischen Beziehung wird ein Effekt nur für Patienten erwartet, die die Beziehung zu ihrem Bezugstherapeuten zu Beginn der Behandlung als unterdurchschnittlich einschätzen.

##### Hypothesen:

Patienten in Gruppe 3, die die therapeutische Beziehung zu Behandlungsbeginn und/oder bei der ersten Zwischenerhebung als unterdurchschnittlich einschätzen, zeigen im Vergleich zu Patienten der Gruppen 1 und 2 mit initial unterdurchschnittlich bewerteter therapeutischer Beziehung:

1.9.4. a) einen höheren (besseren) Wert auf dem „Helping Alliance“-Beziehungsscore bei Entlassung

1.9.4 .b) einen höheren Anteil an reliabler Verbesserung in den „Helping Alliance“-Werten von Aufnahme bis Entlassung

## 2 MATERIAL UND METHODEN

### 2.1 Rahmenbedingungen und Behandlungssetting

Die Daten für diese Arbeit wurden im Zeitraum von Januar 2005 bis November 2007 im Niedersächsischen Landeskrankenhaus (NLKH) Tiefenbrunn erhoben (jetzt: „Asklepios Fachklinikum Tiefenbrunn“). Das NLKH Tiefenbrunn ist ein Krankenhaus für Psychotherapie, Psychiatrie und psychosomatische Medizin und akademisches Lehrkrankenhaus der Georg-August-Universität Göttingen. Die Behandlung ist von der Grundkonzeption her psychodynamisch ausgerichtet. Die Patienten nahmen am multimodalen Behandlungsprogramm der Klinik teil, das schwerpunktmäßig psychodynamisch orientierte und psychoanalytisch-interaktionelle Einzel- und Gruppentherapie, Gestaltungstherapie, ergotherapeutische und sozialtherapeutische Aktivitäten sowie körpertherapeutische Verfahren umfasst. In den ersten vier Behandlungswochen erhielten alle Patienten zwei Einzelgespräche à 45 Minuten mit ihrem Einzeltherapeuten, in den darauffolgenden Wochen ein Einzelgespräch à 45 Minuten pro Woche. Die Häufigkeit des einzeltherapeutischen Kontaktes war für alle Patienten gleich und entsprach der Klinikroutine; Umfang und Dauer der Einzeltherapie wurden für diese Studie nicht modifiziert.

Die Studie wurde auf allen Stationen durchgeführt, die bereits routinemäßig Patientenerhebungen zu Behandlungsbeginn, zu einem Zeitpunkt im Behandlungsverlauf und bei Entlassung im Rahmen der internen Qualitätssicherung implementiert hatten

Die Studie wurde ohne gesonderte finanzielle oder personelle Förderung durchgeführt. Die lokale Studienkoordination hatte die Abteilung Dokumentation und Qualitätssicherung des NLKH Tiefenbrunn inne.

### 2.2 Studiendesign

Es wurde ein prospektives, cluster-randomisiertes Studiendesign gewählt (Abbildung 3). Die Randomisierung erfolgte auf der Ebene der Therapeuten, Untersuchungseinheit war die Ebene der Patienten.

Die Bezugstherapeuten, die in die Studienteilnahme eingewilligt hatten, wurden einer der folgenden drei Bedingungen zugewiesen:

- 1) Symptombezogene Rückmeldung sechs Wochen nach Aufnahme und bei Entlassung („späte“ Rückmeldung, symptombezogen),
- 2) Symptombezogene Rückmeldung bei Aufnahme, zwei Wochen und sechs Wochen nach Aufnahme sowie bei Entlassung („frühe“ und „späte“ Rückmeldung, symptombezogen),
- 3) Sowohl symptom- als auch prozessbezogene Rückmeldung (= Feedback über die therapeutische Arbeitsbeziehung) bei Aufnahme, zwei Wochen und sechs Wochen

nach Aufnahme sowie bei Entlassung („frühe“ und „späte“ Rückmeldung, symptom- und prozessbezogen).

Alle an der Studie teilnehmenden Patienten wurden mit dem identischen Fragebogeninventar und mit gleicher Häufigkeit - bei Aufnahme, zwei Wochen nach Aufnahme, sechs Wochen nach Aufnahme und bei Entlassung - befragt.

Die Patienten waren bezüglich der Feedbackbedingung verblindet. Allerdings stand es den Therapeuten offen, die Rückmeldungen mit den Patienten zu besprechen.

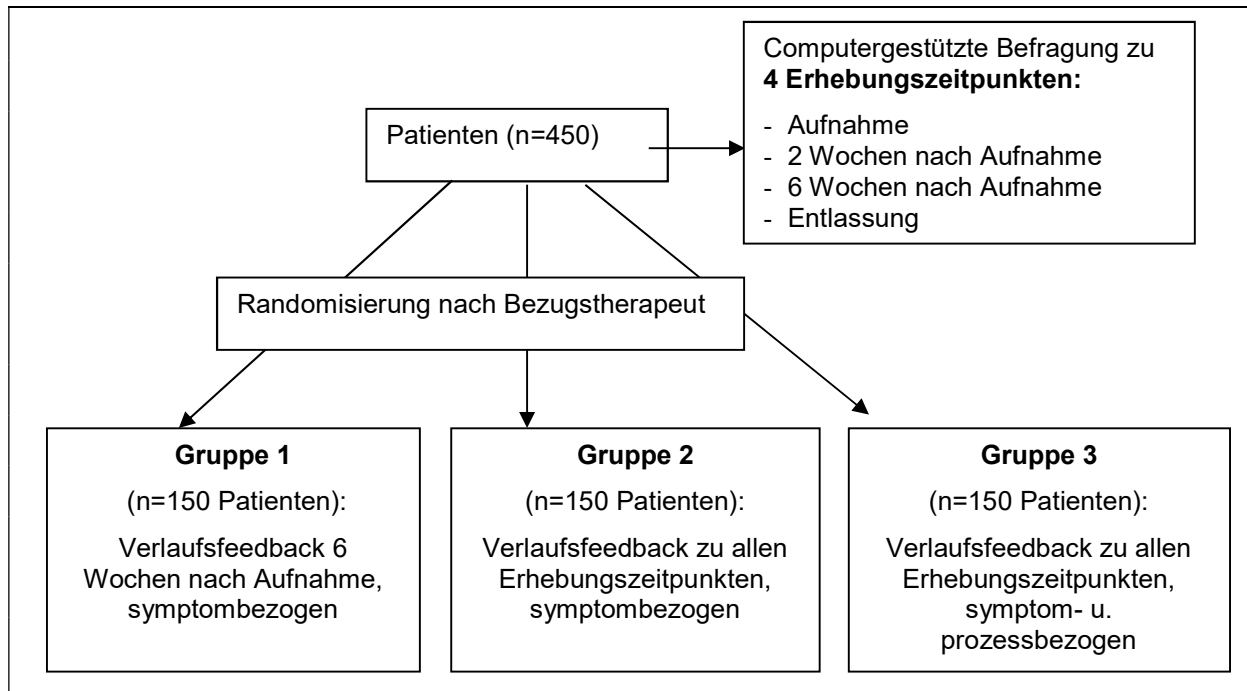


Abbildung 3: Studiendesign

**Stichprobengröße.** Hier wurden drei „aktive“ Gruppen verglichen, die alle das gleiche umfangreiche stationäre Behandlungsprogramm erhielten, und das Verlaufsfeedback in den Gruppen 2 und 3 war lediglich ein „add-on“ zum bereits etablierten Feedback sechs Wochen nach Behandlungsbeginn. Daher wird für das zusätzliche, früh im Behandlungsverlauf bereitgestellte Feedback ein kleiner Effekt erwartet.

### 2.2.1 Ein- und Ausschlusskriterien

Aus organisatorischen Gründen wurden Stationen von der Studienteilnahme ausgeschlossen, auf denen keine routinemäßigen Patientenbefragungen stattfanden. Dies betraf die Gerontopsychiatrie, Stationen mit dem Behandlungsschwerpunkt akute Psychosen sowie Stationen, die Patienten mit Abhängigkeitserkrankungen behandeln. Dieses Patientenkontingent war demnach nicht in der Studienpopulation vorhanden. Aus organisatorischen Gründen wurden außerdem Patienten, bei denen eine Kurzzeittherapie geplant war, nicht in die Studie einbezogen.

Auf den an der Datenerhebung beteiligten Stationen war diese nicht auf Patientengruppen mit bestimmten Störungsbildern beschränkt. Aus der Erfahrung mit Pilotprojekten stellt die computergestützte Dateneingabe auch für Personen, die im Umgang mit dem Computer nicht oder wenig geübt sind, keine Schwierigkeit dar; die Mehrheit der Patienten zieht die Dateneingabe am Computer der Paper-Pencil-Methode vor (Percevic, 2005).

Es wurden Intent-to-treat Analysen mit folgenden Einschlusskriterien durchgeführt:

- Der behandelnde Bezugstherapeut gab sein Einverständnis zur Studienteilnahme.
- Der Patient gab sein schriftliches Einverständnis zur Teilnahme an der Studie.
- Die Behandlung wurde regulär beendet (keine externe Notfall-Verlegung oder Behandlungsabbruch). Dies wurde operationalisiert durch das Vorliegen auswertbarer Patientenfragebögen sowohl bei Aufnahme als auch bei Entlassung.
- Es lagen auswertbare Datensätze bei der ersten Zwischenerhebung vor.

Das letzte Kriterium wurde gewählt, da die Prüfung der zentralen Hypothesen die Berechnung eines Auffälligkeitssignals im frühen Behandlungsverlauf voraussetzt, was nur bei Vorliegen der Patientenfragebögen zur ersten Zwischenerhebung möglich ist.

### 2.3 Studienorganisation und Ablauf der Datenerhebung

Vor Beginn der Studie fand im Rahmen der klinikinternen Fortbildung eine Informationsveranstaltung statt, in der alle Bezugstherapeuten, das Pflegepersonal und die Klinikleitung von der Doktorandin über die Ziele und den Ablauf der Studie informiert wurden. Therapeuten und Pflegepersonal erhielten außerdem eine schriftliche Erläuterung zu den Inhalten des Prozess- und Ergebnisfeedbacks (vgl. Anhang A). Rückmeldungen über die Behandlungsergebnisse der Patienten waren den Therapeuten bereits aus der internen Qualitätssicherung nach dem Stuttgart-Heidelberger Modell der QS vertraut (vgl. 1.2).

Alle Patienten, die für eine Teilnahme infrage kamen (s.o.), wurden zu Behandlungsbeginn von ihren Bezugstherapeuten über die Studie aufgeklärt und gaben, wenn sie zur Teilnahme bereit waren, ihr schriftliches Einverständnis, dass ihre Daten zu wissenschaftlichen Zwecken analysiert werden dürfen.

Die Teilnahme der Therapeuten an der Studie war freiwillig.

Die Patienten wurden bei Aufnahme, wie in der Klinikroutine üblich, denjenigen Therapeuten als Bezugstherapeuten zugewiesen, die zu diesem Zeitpunkt freie Kapazitäten hatten. Dabei werden keine Selektionsmechanismen angewandt.

Die Datenerhebung, -analyse und -rückmeldung erfolgte computergestützt durch das Softwareprogramm Web-AKQUASI (Percevic et al., 2006; Zimmer & Mößner, 2012). Web-AKQUASI wurde von der Forschungsstelle für Psychotherapie entwickelt, um Ergebnismonitoring und Qualitätssicherung in der Psychotherapie zu unterstützen (vgl. 1.2). Alle Patienten des NLKH Tiefenbrunn werden zum Zweck der Qualitätssicherung bei

Aufnahme, sechs Wochen nach Aufnahme (falls keine Kurzzeitbehandlung stattfindet) und bei Entlassung mit einem standardisierten Fragebogeninventar befragt. Für die vorliegende Studie wurde zusätzlich eine Patientenerhebung zwei Wochen nach Aufnahme eingeführt. Alle Patienten füllten die Fragebögen an Computern aus, die auf den Stationen zur Verfügung stehen. Mitarbeiter des Stationsteams, die von der Abteilung Dokumentation und Qualitätssicherung im Umgang mit dem Programm Web-AKQUASI geschult worden waren, wiesen die Patienten in die Dateneingabe ein. Mitarbeiter der Abteilung Dokumentation und Qualitätssicherung standen klinikintern für die Bezugstherapeuten und das Pflegepersonal als Ansprechpartner bei Fragen zur Anwendung von Web-AKQUASI oder zu Inhalten der Feedbacknachrichten zur Verfügung. Dabei erfolgte eine enge Kooperation mit der Forschungsstelle für Psychotherapie, an der die Doktorandin im Studienzeitraum für die Implementierung und Adaptation des Programms verantwortlich war.

Die administrativen Aufgaben im Rahmen der Studie entsprachen im Wesentlichen denen von Web-AKQUASI im Routinebetrieb (z.B. Administration von Therapeuten und Patienten, Einstellen des Erhebungsplans, Organisation der computergestützten Patientenbefragung).

Die unterschiedlichen Rückmeldungen - kein Zugriff auf Rückmeldefunktionen nach Aufnahme und zum frühen Feedbackzeitpunkt für Therapeuten der Gruppe 1, ergebnisbezogenes Feedback für Therapeuten der Gruppen 1 und 2, prozess- und ergebnisbezogenes Feedback für Therapeuten der Gruppe 3 - wurden durch Web-AKQUASI automatisch generiert. Dafür wurde das Programm zu Studienzwecken adaptiert und eine Variable programmiert, durch die die Therapeuten einer der drei Bedingungen zugewiesen wurden.

Um sicher zu stellen, dass die Therapeuten die Rückmeldungen erhalten haben und um deren Akzeptanz und wahrgenommenen Nutzen zu untersuchen, beantworteten die Therapeuten bei jedem Erhebungszeitpunkt, ebenfalls computergestützt, einen kurzen Feedbackbogen (vgl. Anhang B).

## 2.4 Erhebungsinstrumente

### 2.4.1 Inhaltliche und methodische Aspekte

Es besteht Konsens, Therapieerfolg *multidimensional* abzubilden und neben Veränderungen im psychischen Bereich auch die physische und interpersonal-soziale Ebene zu berücksichtigen (z.B. Sperry, Brill, Howard, & Grissom; 1996, Schulte, 1993; Kordy, 1992; Lutz et al., 1996). Da die Beurteilungen des Therapieerfolgs aus unterschiedlichen Perspektiven gering korrelieren (Beutler & Hamblin, 1986; Strupp, Horowitz, & Lambert, 1997), sollten neben mehreren Merkmalsdimensionen auch unterschiedliche Beurteilungsperspektiven berücksichtigt werden; Therapieerfolg soll also sowohl *multidimensional* als auch *multimodal* erfasst werden (Schulte, 1993). Außerdem wird empfohlen, zur Bestimmung des

Therapieerfolgs eine Kombination aus direkter und indirekter Veränderungsmessung einzusetzen (Michalak, Kosfelder, Meyer, & Schulte, 2003), da retrospektive Erfolgsmaße (direkte Veränderungsmessung) und über den prä-post-Vergleich bestimmte Veränderungen (indirekte Veränderungsmessung) gering miteinander korrelieren (z.B. Baumann, Sodemann, & Tobien, 1980; Kastner & Basler, 1997).

Durch eine Kombination aus retrospektiver Erfolgsbeurteilung (direkte Veränderungsmessung) und indirekter prä-post-Veränderungsmessung sowie der Berücksichtigung unterschiedlicher Beurteilungsperspektiven (Patienten- und Therapeutenrating) sollte bei der Erhebung der Daten für die vorliegende Arbeit der Forderung nach einer multimodalen, multidimensionalen und multimethodalen Ergebnismessung (Schulte, 1993) entsprochen werden.

Auf den Einsatz diagnosespezifischer Instrumente wurde verzichtet. Dies geschah einerseits aus Gründen der Praktikabilität: So ermöglicht ein identisches Fragebogeninventar für alle Patienten die Datenerhebung und -rückmeldung an den Bezugstherapeuten unmittelbar nach der Aufnahme – es muss nicht gewartet werden, bis der diagnostische Prozess abgeschlossen ist. Andererseits erschien die Beschränkung auf eine diagnoseübergreifende Erfassung der psychischen Symptomatik auch inhaltlich gerechtfertigt: So finden sich hohe Interkorrelationen zwischen diagnosespezifischen Maßen und den inhaltlich konvergenten Subskalen der SCL-90-R, z.B.  $r=0,81$  zwischen der SCL-90-Skala „Depressivität“ und der Depressivitätsskala (von Zerssen, 1986) oder  $r=0,68$  zwischen der SCL-90-Skala „Somatisierung“ und der Beschwerden-Liste (von Zerssen, 1986).

#### 2.4.2 Patientenfragebögen

In der vorliegenden Arbeit wurden folgende Instrumente verwendet, um die Symptomatik (psychische und interpersonale Beeinträchtigung) und die therapeutische Beziehung aus Patientensicht zu erfassen sowie eine direkte Einschätzung der in der Therapie erzielten Veränderungen zu erhalten.

##### Symptomcheckliste 90-R

Die psychische Beeinträchtigung wurde mit der Symptomcheckliste 90-R (SCL-90-R, Originalfassung: Derogatis, 1986, deutsche Übersetzung: Franke, 1995) erfasst. Die SCL-90-R umfasst 90 Items mit 5-stufigem Antwortformat (0: überhaupt nicht, 4: sehr stark), aus denen folgende neun Subskalen gebildet werden: „Somatisierung“, „Zwanghaftigkeit“, „Unsicherheit im Sozialkontakt“, „Depressivität“, „Ängstlichkeit“, „Aggressivität und Feindseligkeit“, „Phobische Angst“, „paranoides Denken“, „Psychotizismus“. Für den Globalen Schwereindex (GSI), der die Gesamtbeeinträchtigung abbildet, werden über verschiedene Stichproben hinweg hohe interne Konsistenzwerte (Cronbach  $\alpha$  zwischen .94 und .98) und gute Retest-

Reliabilitäten ( $r_{tt}$  zwischen .69 und .92) berichtet (Franke, 1995). Für die meisten Subskalen liegen Befunde vor, die die konvergente und diskriminante Validität stützen (Franke, 1995). Die SCL-90-R bzw. Kurzformen des Verfahrens werden vielfach aufgrund der häufigen Verwendung zum Einsatz bei der Therapieevaluation (z.B. Schulte, 1993; Grawe & Braun, 1994), der Basisdokumentation (Heuft & Senf, 1998; Faller et al., 1997) oder der Qualitätssicherung (u.a. Lutz, 1996; Scheidt et al., 1999) empfohlen.

### Inventar Interpersonaler Probleme

Interpersonelle Beeinträchtigung wurde mit dem Inventar Interpersonaler Probleme (IIP, Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988; deutsche Fassung: Horowitz, Strauß, & Kordy, 2000) erhoben. Das IIP ist ein Selbstbeurteilungsinstrument, das zur Status- und Veränderungsdiagnostik in der klinischen Psychologie, Psychotherapie und Persönlichkeitspsychologie verwendet wird. Es wurde die Kurzform des IIP (64 Items) verwendet. Das IIP basiert auf Circumplex-Modellen der Persönlichkeit, die Persönlichkeitsstile in einem zweidimensionalen interpersonalen Raum mit den beiden Polen „dominant vs. submissiv“ und „freundlich vs. abweisend“ beschreiben. Die im Circumplex-Modell gebildeten Oktanten werden im IIP durch folgende acht Skalen abgebildet, die jeweils acht fünfstufige Items (0: überhaupt nicht, 4: sehr) umfassen: „autokratisch/ dominant“ (PA), „rachsüchtig/ konkurrierend“ (BC), „abweisend/ kalt“ (DE), „introvertiert/ sozial vermeidend“ (FG), „selbstunsicher/ unterwürfig“ (HI), „ausnutzbar/ nachgiebig“ (JK), „fürsorglich/ freundlich“ (LM), „expressiv/ aufdringlich“ (NO). Die interne Konsistenz (Cronbach  $\alpha$ ) von sieben der acht IIP-Skalen ist mittelmäßig bis zufrieden stellend (.47-.64) und mit .36 niedrig für die Skala „fürsorglich/ freundlich“ (LM); die höchste interne Konsistenz weist die Skala „autokratisch/ dominant“ (PA) auf (.64). Es wurden unipsatierte Skalenrohwerte durch Aufsummieren der Itemwerte pro Skala berechnet. Außerdem wurde der Gesamtwert ( $IIP_{ges}$ ) als gemittelte Summe der Skalenwerte nach der Formel:  $IIP_{ges} = (PA + BC + DE + FG + HI + JK + LM + NO)/8$  bestimmt (Horowitz et al., 2000).

### Helping Alliance Questionnaire

Der Helping Alliance Questionnaire (HAQ, Alexander & Luborsky, 1986; deutsche Fassung: Bassler et al., 1995) erfasst mit 11 Items die Qualität der therapeutischen Beziehung, in der amerikanischen Originalversion nur aus Patientensicht, in der deutschen Fassung von existiert auch eine Therapeutenversion mit entsprechend adaptierten Items. Faktorenanalytisch zeigen sich neben der allgemeinen Summenskala (Cronbach  $\alpha = .89$ ) die Subskalen „Beziehungszufriedenheit“ (6 Items, Cronbach  $\alpha = .89$ ) und „Erfolgswilligkeit“ (4 Items, Cronbach  $\alpha = .84$ ). Da die Subskala „Erfolgswilligkeit“ hoch mit Therapieerfolgskriterien (Symptomskalen) korreliert, wurde der Empfehlung der Autoren der deutschen Fassung



gefolgt, zur Prozessevaluation der therapeutischen Beziehung über mehrere Zeitpunkte hinweg die Skala „Beziehungszufriedenheit“ zu verwenden. Das Prozess-Feedback in Gruppe 3 basiert demnach auf folgenden sechs Items der Skala „Beziehungszufriedenheit“:

- „Ich glaube, dass mein Therapeut mir hilft“ (Item 1)
- „Ich habe das Gefühl, mich auf den Therapeuten verlassen zu können“ (Item 6)
- „Ich habe das Gefühl, dass mich der Therapeut versteht“ (Item 7)
- „Ich habe das Gefühl, dass der Therapeut möchte, dass ich meine Ziele erreiche“ (Item 8)
- „Ich habe das Gefühl, dass ich wie auch der Therapeut ernsthaft an einem Strang ziehen“ (Item 9)
- „Ich glaube, dass ich und der Therapeut meine Probleme ähnlich sehen und beurteilen“ (Item 10).

Wie in der englischen Originalversion wurde der HAQ in einer sechsstufigen Skalierung verwendet. Da für den HAQ zum Zeitpunkt der Datenauswertung keine Normwerte verfügbar waren, werden in dieser Arbeit die HAQ-Daten von 348 Patienten, die zwischen dem Jahr 2003 und 2004 im NLKH Tiefenbrunn behandelt worden waren, als Referenzwerte herangezogen (Kennwerte in Anhang A).

#### Veränderungsfragebogen des Erlebens und Verhaltens

Der Veränderungsfragebogen des Erlebens und Verhaltens (VEV, Zielke & Kopf-Mehnert, 1978) ist ein Maß der direkten, retrospektiven Veränderungseinschätzung. Grundlage der Itemauswahl bildet die Veränderungstheorie der Gesprächspsychotherapie unter Einbezug subjektiver Änderungsaussagen von Klienten. Der VEV ist bipolar konzipiert mit den Polen Entspannung/ Gelassenheit/ Optimismus versus Spannung/ Unsicherheit/ Pessimismus. Eine Veränderung in positiver Richtung ist mit größerer Ruhe und Sicherheit, mehr Selbstsicherheit und der Abnahme innerer Spannungen assoziiert. Der negative Pol steht für Spannung, Unsicherheit und Pessimismus im Erleben und Verhalten. Der VEV enthält 42 siebenstufige, unterschiedlich gepolte Items, die die subjektive Differenz zwischen dem momentanen Zustand (Entlassung) und einem zeitlich vorangehenden Zustand (Aufnahmezeitpunkt) einschätzen (z.B. „Ich fühle mich ruhiger und ausgeglichener“, „Meine Schwierigkeiten im Umgang mit anderen Menschen haben zugenommen“). Es wird ein Summenwert aus allen 42 Veränderungsaussagen gebildet, wobei hohe Testwerte positive Veränderungen und niedrige Testwerte negative signalisieren. Der VEV eignet sich sowohl für die Messung unmittelbarer Therapieeffekte als auch für die Erfassung von Langzeiteffekten (Zielke & Kopf-Mehnert, 1978).

### 2.4.3 Therapeutenangaben

**Diagnosen.** Wie im klinischen Alltag üblich, wurden die Diagnosen nicht durch psychodiagnostische Interviews ermittelt. Die Patienten wurden sowohl nach dem Erstgespräch als auch nach Zweitsicht (zwei Wochen nach Aufnahme) nach der Internationalen Klassifikation der Krankheiten ICD-10 (Dilling, Mambour & Schmidt, 1993) diagnostiziert. In dieser Arbeit werden die Hauptdiagnosen, die nach Zweitsicht gestellt wurden, berichtet, da sie aufgrund einer breiteren Informationsbasis als valider angesehen werden.

**Beeinträchtigungsschwerescore.** Die Beeinträchtigung des Patienten aus Sicht des Therapeuten wurde mit Hilfe des Beeinträchtigungsschwerescores (BSS, Schepank, 1995) erfasst. Die Aspekte „körperliche Beeinträchtigung“, „psychische Beeinträchtigung“ und „sozialkommunikative Beeinträchtigung“ werden auf einer fünfstufigen Skala (0: keine Beeinträchtigung, 4: extreme Beeinträchtigung) für den Zeitraum der letzten sieben Tage bewertet. Die Summe der Werte dieser drei Bereiche drückt die Gesamtbeeinträchtigung aus. Nach Schepank (1995) gilt ein Patient dann als „Fall“, wenn der Schweregrad seiner Gesamtbeeinträchtigung 5 und mehr Punkte erreicht. Von einer sehr gravierenden Gesamtbeeinträchtigung wird bei einem Summenwert von mehr als 8 Punkten ausgegangen.

**Fragebogen zur Beurteilung der Rückmeldungen (Feedbackbogen).** Um die Akzeptanz der Rückmeldungen zu untersuchen, wurde ein kurzer Feedbackbogen (ca. drei Minuten Bearbeitungszeit) entwickelt, der vom behandelnden Therapeuten nach jeder Rückmeldung auszufüllen war. Der Feedbackbogen wurde ebenfalls in Web-AKQUASI computergestützt vorgegeben. Gleichzeitig diente dieser Bogen zur Überprüfung, ob die Behandler die Rückmeldungen gesehen hatten.

Der Feedbackbogen erfasste a) die Übereinstimmung der Rückmeldung mit der eigenen Einschätzung, b) die Relevanz der enthaltenen Informationen, c) die Beurteilung des Nutzens der Rückmeldung, und d) Änderungen von Behandlungsmaßnahmen als Reaktion auf die Rückmeldung.

Eine Übersicht über das Fragebogeninventar für Patienten und Therapeuten zu den vier Erhebungszeitpunkten ist Tabelle 1 zu entnehmen. Bei Aufnahme und/oder Entlassung wurden weitere Informationen und Instrumente erhoben (z.B. zur Therapiezielerreichung oder Patientenzufriedenheit), die jedoch nicht für die Analysen herangezogen wurden und auf die daher nicht weiter eingegangen wird.

Tabelle 1: Fragebogeninventar für Patienten und Therapeuten zu den vier Erhebungszeitpunkten

Erhebungszeitpunkt	Fragebögen	
	Patient	Therapeut
<b>Aufnahme</b>	SCL-90-R IIP HAQ LQ	BSS HAQ Feedbackbogen
<b>1. Zwischen- erhebung</b>	SCL 90-R HAQ	Feedbackbogen
<b>2. Zwischen- erhebung</b>	SCL 90-R IIP HAQ	HAQ Feedbackbogen
<b>Entlassung</b>	SCL-90-R IIP HAQ LQ VEV	BSS HAQ Feedbackbogen

*Anmerkungen:* SCL-90-R: Symptom Checkliste 90-Revised; IIP: Inventar Interpersonaler Probleme, HAQ: Helping Alliance Questionnaire; LQ: Fragebogen zur Lebensqualität; VEV: Veränderungsfragebogen des Erlebens und Verhaltens; BSS: Beeinträchtigungsschwerescore

## 2.5 Inhalte der Rückmeldungen

In Web-AKQUASI erfolgen Datenanalyse und –aufbereitung unmittelbar nach Ausfüllen der Fragebögen. Die Informationen stehen demnach ohne Zeitverzögerung zur Verfügung.

Aus den Patientenfragebögen wurden folgende Informationen zurück gemeldet:

**Aktuelle Beeinträchtigung.** Die Subskalenwerte des IIP und der SCL-90-R sowie der GSI als Maß der globalen psychischen Beeinträchtigung wurden als Perzentilwerte in Relation zu klinisch unauffälligen Referenzpopulationen dargestellt. Perzentilwerte wurden gewählt, um eine Vergleichbarkeit über verschiedene Instrumente hinweg zu gewährleisten und die Interpretation der Skalenwerte zu erleichtern. Als Cut-Off zwischen funktionalem und dysfunktionalem Bereich wurde das 84. Perzentil gewählt.

**Veränderung.** Ab dem zweiten Messzeitpunkt (= erste Zwischenerhebung) wurde die Veränderung auf den IIP- und SCL-Skalen sowie für Gruppe 3 auf der HAQ-Beziehungsskala nach dem Prinzip der reliablen und klinisch signifikanten Veränderung dargestellt (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991).

**Zusammenfassung und Bewertung.** Jede Rückmeldung enthält eine Kurzzusammenfassung und Bewertung des Behandlungsverlaufs in Textform. Diese bezieht sich auf die aktuelle Symptomatik (Zielbereich unterhalb des cut-offs erreicht oder nicht erreicht) und den Symptomverlauf (überwiegend Verbesserung, keine Veränderung, überwiegend Verschlechterung, ähnliches Ausmaß an Verbesserung und Verschlechterung). In Gruppe 3 erfolgte zusätzlich eine Bewertung der therapeutischen Beziehung aus Patientensicht (Therapeutische Beziehung positiv oder unterdurchschnittlich, Veränderung der therapeutischen Beziehung im Therapieverlauf).







Wenn ein Verlauf nach den in Abbildung 5 aufgeführten Kriterien (vgl. 2.6.1) als „auffällig“ bewertet wurde, wurde das Feedback zusätzlich mit der Textmeldung „auffälliger Verlauf“ versehen sowie mit den Schlüsselitems und/oder Skalenbewertungen, die das Auffälligkeitssignal ausgelöst hatten, z.B. „auffälliger Verlauf [mindestens genauso viele negative wie positive Veränderungen, starke Suizidgedanken]“.

Abbildung 4 zeigt ein Beispiel für eine Feedbacknachricht zum Zeitpunkt der ersten Zwischenerhebung (zwei Wochen nach Aufnahme) für Gruppe 3, in der sowohl Rückmeldung über die therapeutische Beziehung aus Patientensicht (erste Zeile der Ergebnistabelle) als auch bezüglich der psychischen Symptomatik (Subskalen der SCL-90-R sowie Allgemeiner Symptomindex als Gesamtwert) erfolgte.

		Bereich	Zustand	Veränd.
☑	☑	therap. Beziehung (Pat.)	30	↔
☑	☑	Somatisierung	97	↓
☑	☑	Zwanghaftigkeit	88	→
☑	☑	Unsicherheit im Sozialkontakt	54	↑
☑	☑	Depressivität	66	↑
☑	☑	Ängstlichkeit	76	↑
☑	☑	Aggressivität und Feindseligkeit	95	↓
☑	☑	Phobische Angst	99	→
☑	☑	Paranoides Denken	96	→
☑	☑	Psychotizismus	86	↗
☑	☑	Allgemeiner Symptom Index	94	↗

Abbildung 4: Beispiel für eine Rückmeldung in Gruppe 3 (prozess- und symptombezogenes Feedback)

Die Pfeile in der Spalte „Veränd“ (Veränderung) haben folgende Bedeutung:

-  keine Veränderung; Skalenwert im Zielbereich (< 84. Perzentil)
-  keine Veränderung; Skalenwert nicht im Zielbereich (> 84. Perzentil)
-  reliable Verbesserung
-  klinisch signifikante Verbesserung
-  reliable Verschlechterung
-  klinisch signifikante Verschlechterung

### 2.5.1 Bestimmung von reliabler und klinisch signifikanter Veränderung

Reliable und klinisch signifikante Veränderungen wurden nach dem Konzept von Jacobson und Truax (1991) bestimmt, wonach der Unterschied in den Werten zweier Messzeitpunkte als klinisch bedeutsam einzustufen ist, wenn a) die Differenz der Werte größer ist als der Messfehler des Instruments (statistisch signifikante bzw. „reliable“ Veränderung) und b) der zuletzt gemessene Wert den Werten einer „gesunden“ Normpopulation ähnlicher ist als Werten, die typischerweise in Patientenpopulationen gefunden werden. Das Vorgehen nach Jacobson und Truax (1991) liefert, verglichen mit anderen Methoden zur Bestimmung klinischer Signifikanz, eine moderate (weder übermäßig optimistisch noch pessimistische) Schätzung klinisch signifikanter Veränderung (Bauer, Lambert, & Nielsen, 2004). Als Cut-off Score zwischen dem funktionalen („gesunden“) und dem dysfunktionalen („krankheitswertigen“) Bereich wurde das 84. Perzentil der Bevölkerungsstichprobe verwendet (Percevic, Bauer, & Kordy, 2004).

Eine reliable Veränderung wurde bei Zugrundelegung eines 95%-Konfidenzintervalls angenommen, wenn der Veränderungswert nach dem Reliable Change Index (RCI) > 1.96 (reliable Zunahme des Messwerts) bzw. < -1.96 (reliable Abnahme des Messwerts) war.

Der RCI berechnet sich aus der Differenz zweier Testwerte einer Person, geteilt durch den Standardfehler der Differenz, und wurde nach folgender Formel bestimmt:

$$RCI = \frac{x_{post} - x_{prä}}{\sqrt{2(s\sqrt{1-r})^2}}$$

Anmerkungen:  $x_{post}$ : Wert bei Therapieende (bzw. der zuletzt gemessene Wert),  $x_{prä}$ : Wert bei Therapiebeginn (bzw. der zuerst gemessene Wert),  $s$ : Standardabweichung,  $r$ : Reliabilität des Messinstruments.

Als Maß für die Reliabilität wurde die interne Konsistenz der jeweiligen Skala (Cronbach Alpha) verwendet. Für die SCL-90- und IIP-Skalen wurden die Standardabweichungen aus Bevölkerungsnormdaten herangezogen; für die HAQ Beziehungsskala wurden Standardabweichung und interne Konsistenz aus den Daten der klinischen Referenzstichprobe berechnet (Anhang B). Vor Beginn der Studie wurde der HAQ sechs Wochen nach Aufnahme und bei Entlassung erhoben. Da diese Studie auf die Rückmeldungen im Therapieverlauf fokussiert, wurden die HAQ-Daten zum Zeitpunkt „sechs Wochen nach Aufnahme“ für die Berechnung der Referenzwerte verwendet.

### 2.5.2 Gruppenunterschiede bezüglich der Rückmeldungen

**Gruppe 1 und Gruppe 2:** Hier wurde eine komprimierte Zusammenfassung von Behandlungsstatus und -verlauf, basierend auf den Skalen der SCL-90 R und des IIP, gegeben. Das Feedback enthält eine Kurzbewertung (z.B. „unauffälliger Verlauf“, „überwiegend Verbesserung“), den aktuellen Zustand (in Perzentilen bezogen auf eine Bevölkerungsnorm) sowie die Veränderungen seit dem letzten Messzeitpunkt und die Veränderung seit der Aufnahmeerhebung.

**Gruppe 3:** Hier wurde zusätzlich zu dem symptombezogenen Feedback (Gruppe 1 und 2) die therapeutische Arbeitsbeziehung als Prozessmaß zurückgemeldet, wofür die Skala „Beziehungszufriedenheit“ des HAQ verwendet wurde. Bei jedem Erhebungszeitpunkt wird der Skalenwert als Perzentilwert dargestellt. Die therapeutische Arbeitsbeziehung wurde auch in die Kurzbewertung aufgenommen („therapeutische Beziehung positiv“ vs. „therapeutische Beziehung unterdurchschnittlich“) und deren Veränderung im Therapieverlauf zurückgemeldet.

*Tabelle 2:* Instrumente, deren Skalenwerte in den drei Untersuchungsgruppen zu den vier Erhebungszeitpunkten an die Behandler zurückgemeldet werden

Erhebungszeitpunkt	Therapeutenfeedback		
	Gruppe 1	Gruppe 2	Gruppe 3
<b>Aufnahme</b>	<i>Kein Feedback</i>	SCL 90-R	<b>HAQ</b> SCL 90-R
		IIP	IIP
<b>2 Wochen nach Aufnahme</b>	<i>Kein Feedback</i>	SCL 90-R	<b>HAQ</b> SCL 90-R
<b>6 Wochen nach Aufnahme</b>		SCL 90-R IIP	<b>HAQ</b> SCL 90-R IIP
<b>Entlassung</b>		HAQ SCL 90-R IIP	

## 2.6 Operationalisierung von Therapieerfolg

Zur Bestimmung des Therapie-Ergebnisses werden das Ausmaß an psychischer Beeinträchtigung bei Therapieende (Globaler Schwereindex GSI der SCL-90-R) sowie die Raten an reliabler Veränderung bezüglich des GSI (vgl. 2.5.1) betrachtet. Als zentrales Outcome-Maß wird das Auffälligkeitssignal, ermittelt nach dem Algorithmus des Stuttgart-Heidelberger Modells (vgl. 1.2) verwendet.

### 2.6.1 Das Auffälligkeitssignal als aggregiertes Outcome-Maß

In Anlehnung an das Stuttgart-Heidelberger Modell der Qualitätssicherung in der Psychotherapie (vgl. 1.2) wird das Behandlungsergebnis nach einem Algorithmus, der auf qualitativen und quantitativen Aspekten beruht, als „auffällig“ oder „positiv“ bewertet. Bei der Bestimmung des Auffälligkeitssignals wird sowohl die indirekte Veränderungsmessung als auch die direkte Veränderungseinschätzung berücksichtigt; es werden Patienten- und Therapeutenangaben einbezogen. Da dieses Signal die verschiedene Erhebungsperspektiven und –dimensionen integriert und eine valide Ergebnisbewertung psychotherapeutischer Behandlungen erlaubt (vgl. 1.2), wird der Prozentsatz an auffälligen Verläufen in dieser Arbeit als das zentrale Outcome-Maß verwendet.

Wie in Abbildung 5 dargestellt, wird ein Auffälligkeitssignal entweder durch eine „Schlüselfrage“ („Keyitem“) ausgelöst (z.B. der Frage nach Suizidgedanken oder nach einer Veränderung des Allgemeinbefindens, direkte Veränderungseinschätzung) oder durch eine indirekte Veränderungsbewertung, indem die Veränderung von Skalenwerten zwischen zwei Messzeitpunkten nach dem Prinzip der reliablen und klinisch signifikanten Veränderung bewertet wird.

Während die Kriterien für die Definition des Auffälligkeitssignals bei Entlassung der Konzeption des Stuttgart-Heidelberger Modells entsprechen, wurden für die Definition eines „auffälligen“ Zwischenergebnisses Befunde der Therapieverlaufsforschung (vgl. 1.4.3) zugrunde gelegt. Bei Aufnahme wurde ein Fall nur dann als „auffällig“ klassifiziert, wenn der Patient starke oder sehr starke Suizidgedanken äußerte. Bei den Zwischenerhebungen wurden zusätzlich die Veränderungen seit Therapiebeginn berücksichtigt: das Auffälligkeitssignal wurde dann erzeugt, wenn auf den Symptomskalen mehr negative als positive Veränderungen zu beobachten waren oder ein beträchtlicher Anteil der Symptomskalen (mind. 30%) negative Veränderungen aufwies. Im Unterschied zu Feedbacksystemen, die den individuellen Verlauf mit erwarteten Verlaufskurven abgleichen und – basierend auf dem Dosis-Wirkungs-Modell – ein Signal senden, wenn ein Patient keine (hinreichende) Verbesserung im Therapieverlauf zeigt (vgl. 1.4.2, 1.5.1), wurde hier darauf verzichtet, einen Fall während der Therapie bei ausbleibenden Verbesserungen als „auffällig“ zu definieren. Grund hierfür sind die im Rahmen des Random Walk Modells (vgl. 1.4.3) dargestellten Befunde, wonach die Chancen auf

Symptombesserung im Therapieverlauf nicht geringer werden, vorausgesetzt, der Patient wird lange genug behandelt, was im hier untersuchten Setting (stationäre Langzeittherapie) mit höherer Wahrscheinlichkeit gewährleistet werden kann als in den meisten bisher untersuchten Settings.

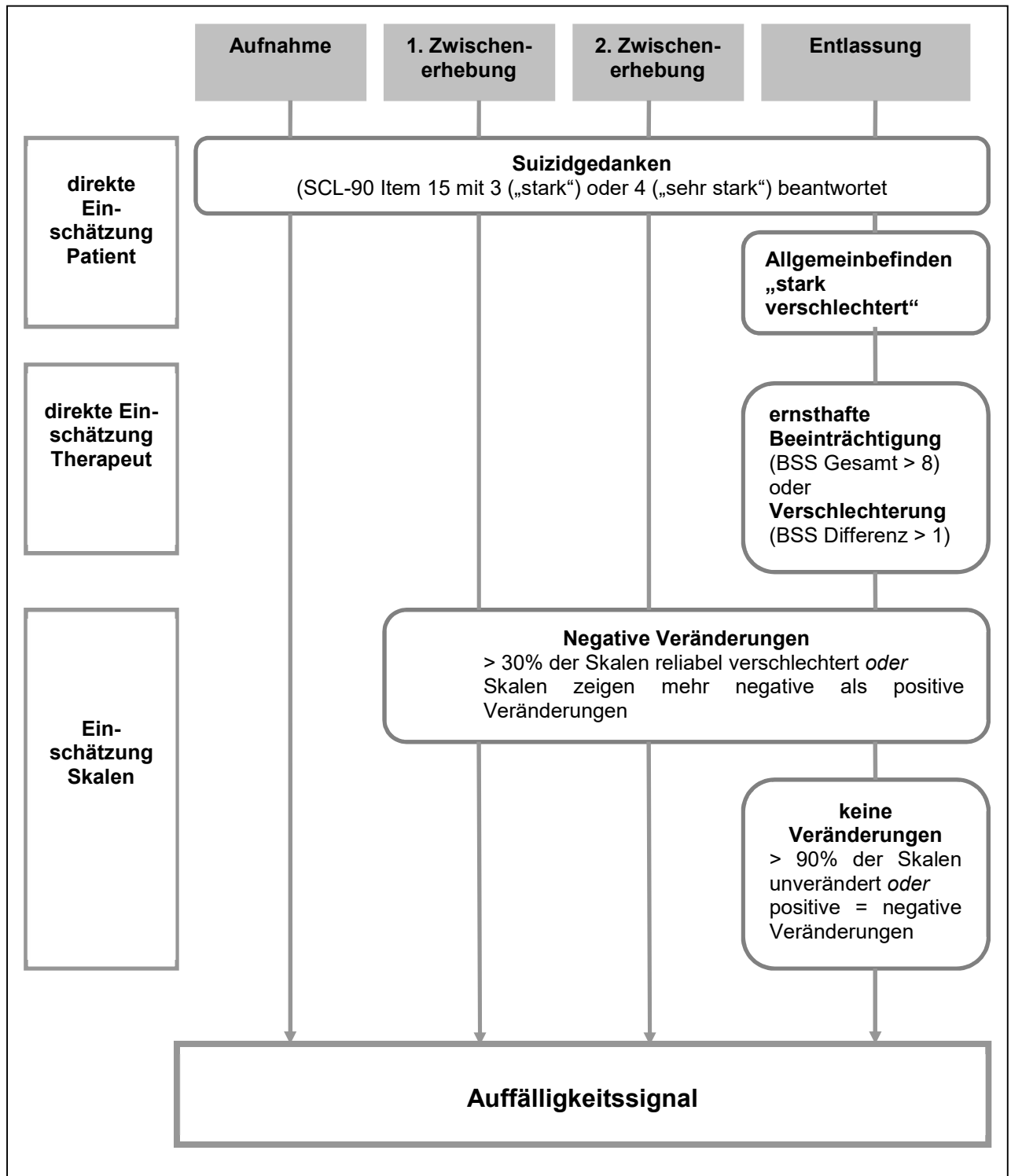


Abbildung 5: Kriterien für die Bewertung eines Behandlungsverlaufs als „auffällig“ zu den vier Erhebungs- und Feedbackzeitpunkten (Anmerkungen: BSS: Beeinträchtigungsschwerescore, BSS Differenz: Differenz des BSS-Gesamtwerts zwischen Aufnahme und Entlassung)



## 2.7 Auswertung

**Stichprobenunterschiede.** Gruppenunterschiede in soziodemografischen und klinischen Variablen werden bei metrischen Variablen durch einfaktorische Varianzanalysen (ANOVA) überprüft. Bei Abweichungen von der Normalverteilung werden Mann-Whitney-U-Tests angewandt. Unterschiede in der Verteilung nominalskalierten Variablen werden mit Chi-Quadrat-Tests überprüft.

**Fragestellung 1.** Gruppenunterschiede in den Therapeutenangaben zu Akzeptanz, Relevanz und wahrgenommenem Nutzen des Feedbacks werden mit Chi-Quadrat-Tests überprüft.

**Fragestellungen 2 bis 4.** Da die Daten hierarchisch organisiert sind (Patienten innerhalb von Therapeuten) werden Mehrebenenmodelle verwendet, um Gruppenunterschiede bezüglich des Therapie-Outcomes (Fragestellung 2), der Behandlungsdauer (Fragestellung 3) und der therapeutischen Beziehung (Fragestellung 4) zu prüfen.

Je nach Skalierung der abhängigen Variable werden gemischte lineare Modelle (bei metrischen abhängigen Variablen, Analysen mit SPSS) oder gemischte logistische Regressionsmodelle (bei binären abhängigen Variablen, Analysen mit SAS) geschätzt.

In allen Modellen ist „Patient“ auf Level 1 und „Therapeut“ auf Level 2 gesetzt. „Therapeut“ wird als zufälliger Faktor, die Variable „Gruppe“ als fester Faktor definiert.

Die Modellparameter werden in den gemischten linearen Modellen mit Hilfe des Restricted Maximum-Likelihood-Verfahrens (REML) und in den gemischten logistischen Regressionsmodellen (GLIMMIX) durch das residual lock pseudo-likelihood-Verfahren (RSPL) geschätzt. Zur Signifikanzprüfung werden F-Tests verwendet. Bei Einzelvergleichen zwischen den drei Gruppen wird bei gemischten linearen Modellen (SPSS) die Bonferroni-Korrektur angewendet, bei gemischten logistischen Modellen (SAS) die Bonferroni-Holm-Korrektur.

**Unabhängige und abhängige Variablen.** Fragestellung 2 (Gruppenunterschiede im Therapie-Outcome) wird durch ein gemischtes lineares Modell (abhängige Variable: GSI bei Entlassung) und zwei gemischte logistische Regressionsmodelle (abhängige Variablen: reliable Veränderung bei Entlassung, Auffälligkeitssignal bei Entlassung) überprüft. Als Prädiktoren werden die Variablen „Gruppe“, „Auffälligkeitssignal bei 1. Zwischenerhebung“ sowie entsprechend der Hypothese die Interaktion „Gruppe\*Auffälligkeitssignal bei 1. Zwischenerhebung“ ins Modell genommen. Als Kovariaten werden die Aufenthaltsdauer und der GSI bei Aufnahme berücksichtigt.

Fragestellung 3 (Gruppenunterschiede bezüglich der Behandlungsdauer) wird durch zwei gemischte lineare Modelle (abhängige Variable: Behandlungsdauer in Tagen) geprüft. Zur Prüfung von Hypothese 3 a) werden die Variablen „Gruppe“, „Auffälligkeitssignal bei 1. Zwischenerhebung“ und die Interaktion „Gruppe\*Auffälligkeitssignal bei 1. Zwischenerhebung“ als Prädiktoren verwendet; als Kovariate wird der GSI bei Aufnahme berücksichtigt. Zur

Prüfung von Hypothese 3 b) werden die Variablen „Gruppe“, „Psychische Beeinträchtigung bei 1. Zwischenerhebung im Normbereich“ und die Interaktion „Gruppe\* Psychische Beeinträchtigung bei 1. Zwischenerhebung im Normbereich“ verwendet; als Kovariate wird ebenfalls der GSI bei Aufnahme berücksichtigt.

Fragestellung 4 (Gruppenunterschiede in der therapeutischen Beziehung) wird durch ein gemischtes lineares Modell (abhängige Variable: Beziehungsskala HAQ bei Entlassung) und ein gemischtes logistisches Regressionsmodell (abhängige Variable: reliable Veränderung der HAQ Beziehungsskala) geprüft. Als Prädiktoren werden die Variablen „Gruppe“, „unterdurchschnittlicher HAQ-Wert bei 1. Zwischenerhebung“ sowie entsprechend der Hypothese die Interaktion „Gruppe\* unterdurchschnittlicher HAQ-Wert bei 1. Zwischenerhebung“ ins Modell genommen. Als Kovariaten werden der GSI bei Aufnahme und die Aufenthaltsdauer berücksichtigt.

### 3 ERGEBNISSE

Es werden zunächst die Therapeuten- und Patientenstichproben beschrieben, daran anschließend erfolgt die Darstellung der Überprüfung der vier Fragestellungen.

#### 3.1 Stichprobenbeschreibung

##### 3.1.1 Therapeutenstichprobe

An der Studie nahmen 21 Einzeltherapeuten – 15 Männer und sechs Frauen - teil, von denen jeweils sieben einer der drei Gruppen zugewiesen wurden. Von jedem Therapeuten konnten im Durchschnitt die Daten von 29,5 Patienten ( $sd=17,5$ ) in die Auswertung einbezogen werden. Dabei variierte die „Caseload“, d.h. die Anzahl der Patienten, die ein Therapeut in die Studie einbrachte, stark: Die Spannweite betrug zwischen 2 und 65 Patienten. Kein Unterschied zeigte sich in der durchschnittlichen „Caseload“ pro Therapeut zwischen den drei untersuchten Gruppen: Therapeuten in Gruppe 1 hatten im Durchschnitt 27,9 ( $sd=17,7$ ) Patienten in der Studie, Therapeuten in Gruppe 2 durchschnittlich 26,7 ( $sd=21,7$ ) Patienten und Therapeuten in Gruppe 3 im Durchschnitt 33,9 ( $sd=14,0$ ) Patienten.

##### 3.1.2 Patientenfluss

Nach dem Intent-to-treat Prinzip gingen Patientendaten in die Datenanalyse ein, wenn jede der folgenden Bedingungen erfüllt war:

- Die Bezugstherapeuten der Patienten gaben ihr Einverständnis zur Studienteilnahme
- Die Patienten gaben ihr schriftliches Einverständnis zur Studienteilnahme
- Es lagen Datensätze zum Aufnahme- und Entlassungszeitpunkt vor
- Patientendaten zur ersten Zwischenerhebung lagen vor (Diese Bedingung war notwendig, um die unter 1.9.2, 1.9.3 und 1.9.4 formulierten zentralen Hypothesen zu überprüfen).

Abbildung 6 zeigt den Patientenfluss, der in der ITT- Stichprobe resultierte.

Alle Bezugstherapeuten der Stationen, auf denen die Studie durchgeführt wurde, willigten in die Teilnahme ein. Im Erhebungszeitraum lag das schriftliche Einverständnis zur Studienteilnahme von 816 Patienten vor. Es fehlten insgesamt 11 Patientendatensätze bei Aufnahme ( $n=1$  in Gruppe 1,  $n=7$  in Gruppe 2 und  $n=3$  in Gruppe 3) und 186 Patientendatensätze bei Entlassung ( $n=62$  in Gruppe 1,  $n=60$  in Gruppe 2 und  $n=75$  in Gruppe 3). Von 619 Patienten mit kompletten Datensätzen bei Aufnahme und Entlassung konnten nur diejenigen Fälle in die Analysen einbezogen werden, bei denen Daten zur ersten Zwischenerhebung vorlagen. Aufgrund fehlender Werte der ersten Zwischenerhebung mussten insgesamt 82 Fälle ausgeschlossen werden ( $n=32$  in Gruppe 1,  $n=19$  in Gruppe 2 und  $n=31$  in Gruppe 3). Es resultierten Daten von 537 Patienten, die in die ITT-Analyse

einbezogen werden konnten. Die ITT-Stichprobe umfasst 163 Patienten in Gruppe 1 (symptombezogenes Verlaufsfeedback 6 Wochen nach Aufnahme), 168 Patienten in Gruppe 2 (symptombezogenes Verlaufsfeedback bei Aufnahme, 2 Wochen und 6 Wochen nach Aufnahme) und 206 Patienten in Gruppe 3 (symptom- und prozessbezogenes Verlaufsfeedback bei Aufnahme, 2 Wochen und 6 Wochen nach Aufnahme).

Die Analysen wurden ausschließlich für die ITT-Stichprobe durchgeführt. Bei Anwendung des per-protocol-Kriteriums hätten lediglich 40 Patienten ausgeschlossen werden müssen. Eine protokollkonforme Durchführung der Studie konnte nicht umgesetzt werden, wenn ein Therapeutenwechsel (z.B. aus organisatorischen Gründen wie Krankheit, Urlaub, interner Verlegung) auch zu einem Wechsel der Untersuchungsgruppe geführt hatte oder die zweite Zwischenerhebung fehlte.

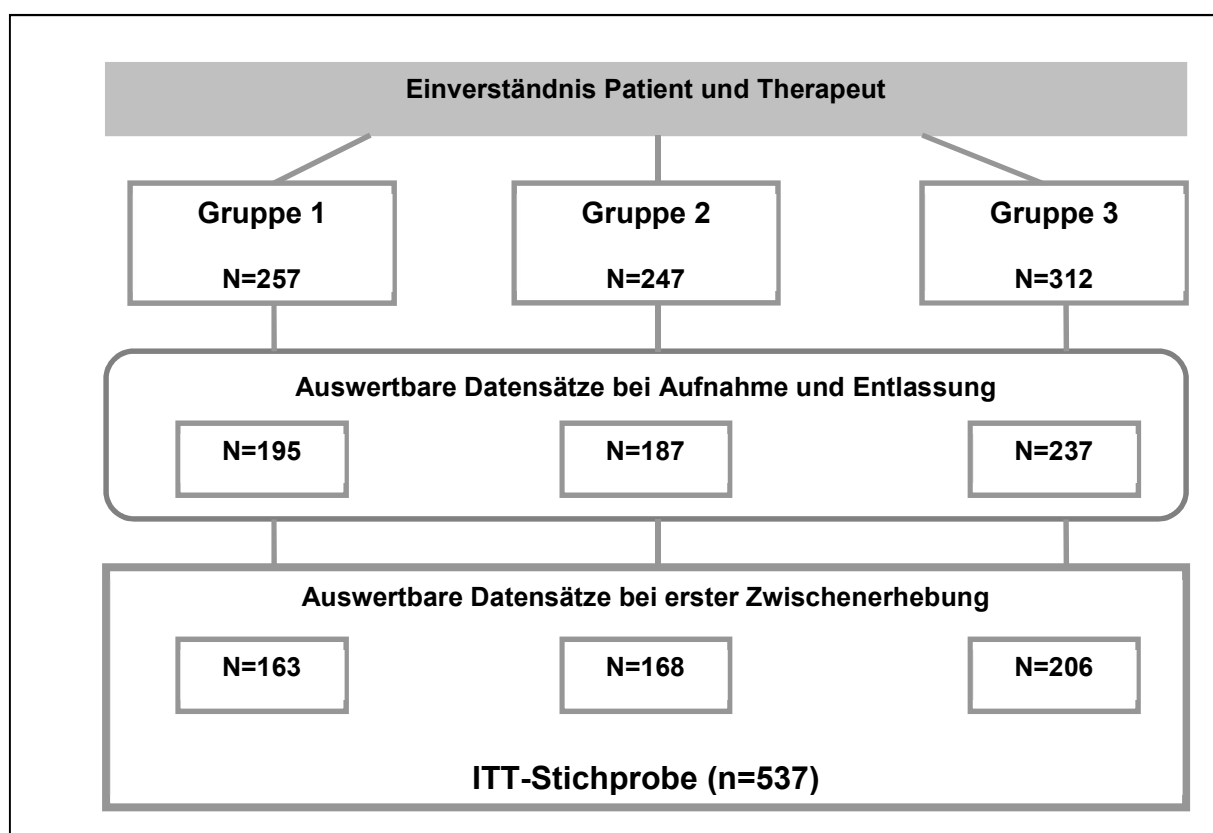


Abbildung 6: Patientenfluss

### 3.1.3 Soziodemografische Merkmale der Patientenstichprobe

In Tabelle 3 sind die soziodemografischen Merkmale der Stichprobe dargestellt. Aufgrund der geringen Häufigkeit von „verwitwet“ (je n=2 in den Gruppen 1 und 2, n=1 in Gruppe 3) wurden geschiedene, getrenntlebende und verwitwete Patienten zu einer Kategorie zusammengefasst.

Tabelle 3: Soziodemografische Merkmale der Patientenstichprobe

	<b>Gruppe 1</b> <b>(n=163)</b>	<b>Gruppe 2</b> <b>(n=168)</b>	<b>Gruppe 3</b> <b>(n=206)</b>	<b>Gesamt</b> <b>(n=537)</b>	<b>Gruppe 1 vs.</b> <b>Gruppe 2 vs.</b> <b>Gruppe 3</b>
<i>Alter</i>					
M (sd)	33.56 (10.87)	33.79 (11.08)	36.07 (11.02)	34.60 (11.04)	F=3.03 (df=2) p=.049
<i>Geschlecht</i>					
Männlich, n (%)	61 (37.4)	56 (33.3)	69 (33.5)	186 (34.6)	X <sup>2</sup> =0.804 (df=2), p=.669
Weiblich, n (%)	102 (62.6)	112 (66.7)	137 (66.5)	351 (65.4)	
<i>Familienstand</i>					
Ledig, n (%)	110 (67.5)	111 (66.1)	121 (58.7)	342 (63.7)	X <sup>2</sup> =5.697 (df=4), p=.223
Verheiratet, n (%)	24 (14.7)	29 (17.3)	49 (23.8)	102 (19.0)	
Getrennt/geschieden/ verwitwet, n (%)	29 (17.8)	28 (16.7)	36 (17.5)	93 (17.3)	
<i>Feste Partnerschaft, n (%)</i>	55 (33.7)	65 (38.7)	102 (49.5)	222 (41.3)	
<i>Schulabschluss</i>					
ohne, n (%)	2 (1.2)	5 (3.0)	1 (0.5)	8 (1.5)	X <sup>2</sup> =17.112 (df=8), p=.029
Hauptschule, n (%)	29 (17.8)	30 (17.9)	32 (15.5)	91 (16.9)	
Realschule, n (%)	63 (38.7)	46 (27.4)	58 (28.2)	167 (31.1)	
(Fach)Abitur, n (%)	65 (39.9)	74 (44.0)	107 (51.9)	246 (45.8)	
Sonstiges, n (%)	4 (2.5)	13 (7.7)	8 (3.9)	25 (4.7)	
<i>Berufsabschluss</i>					
ohne, n (%)	38 (23.3)	38 (22.6)	29 (14.1)	105 (19.6)	X <sup>2</sup> =19.598 (df=10), p=.033
In Ausbildung, n (%)	18 (11.0)	20 (11.9)	26 (12.6)	64 (11.9)	
Lehre, n (%)	53 (32.5)	61 (36.3)	63 (30.6)	177 (33.0)	
Meister/Fachschule, n (%)	17 (10.4)	11 (6.5)	25 (12.1)	53 (9.9)	
FH/ Universität, n (%)	16 (9.8)	21 (12.5)	44 (21.4)	81 (15.1)	
Sonstiges, n (%)	21 (12.9)	17 (10.1)	19 (9.2)	57 (10.6)	
<i>Beschäftigung vor Aufnahme</i>					
voll erwerbstätig, n (%)	49 (30.1)	42 (25.0)	71 (34.5)	162 (30.2)	X <sup>2</sup> =11.920 (df=6), p=.064
Teilzeit beschäftigt, n (%)	22 (13.5)	23 (13.7)	40 (19.4)	85 (15.8)	
nicht erwerbstätig, n (%)	46 (28.2)	58 (34.5)	58 (28.2)	162 (30.2)	X <sup>2</sup> =3.581 (df=2), p=.167
arbeitslos, n (%)	46 (28.2)	45 (26.8)	37 (18.0)	128 (23.8)	
<i>vorzeitige Berentung, n (%)</i>	27 (16.6)	18 (10.7)	22 (10.7)	67 (12.5)	

Ebenfalls aufgrund der geringen Zellenbesetzung wurden bei der Variablen „höchster Schulabschluss“ die Kategorien „Sonderschule“ (keine einzige Nennung), „noch in der Schule“

(n=7 in der Gesamtstichprobe) und „sonstiger Schulabschluss“ (n=18 in der Gesamtstichprobe) zur Kategorie „Sonstiges“ zusammengefasst.

Keine Unterschiede zwischen den drei Gruppen zeigten sich in den Variablen Geschlecht – in allen Gruppen waren ca. zwei Drittel der Patienten weiblich –, Familienstand, Beschäftigungssituation vor Aufnahme und der Häufigkeit einer vorzeitigen Berentung. Dagegen unterschieden sich die Untersuchungsgruppen in den Variablen Alter, Partnerschaft, höchster Schulabschluss und höchster Berufsabschluss. Patienten der Gruppe 3 waren mit einem Durchschnittsalter von 36,1 Jahren signifikant älter als Patienten der Gruppen 1 und 2. Knapp die Hälfte der Patienten in Gruppe 3, aber nur etwa ein Drittel der Patienten in den Gruppen 1 und 2 lebte in fester Partnerschaft. In der Schulbildung zeigten sich Unterschiede zwischen den Gruppen 1 und 3: In Gruppe 3 war der Anteil an Patienten mit (Fach-)Abitur signifikant höher und der Anteil an Patienten mit Realschulabschluss signifikant niedriger als in Gruppe 1. Auch das formale Berufsausbildungsniveau war in Gruppe 3 höher: Patienten in Gruppe 3 gaben in etwa doppelt so häufig einen Hochschulabschluss als höchsten Abschluss an als Patienten der Gruppen 1 und 2. Der Anteil an Personen ohne Berufsabschluss war in den Gruppen 1 und 2 signifikant höher als in Gruppe 3.

#### 3.1.4 Klinische Merkmale der Patientenstichprobe

In Tabelle 4 sind die klinischen Merkmale der drei Untersuchungsgruppen sowie der Gesamtstichprobe zum Aufnahmezeitpunkt aufgeführt. Die große Mehrheit der Patienten erhielt die Erstdiagnose „Affektive Störung“, gefolgt von der Kategorie „Neurotische, Belastungs- und somatoforme Störungen“. In Gruppe 3 erhielten signifikant mehr Patienten eine Erstdiagnose aus der Kategorie „Neurotische, Belastungs- und somatoforme Störungen“ als in den Gruppen 1 und 2, dafür wurden in Gruppe 3 signifikant weniger Patienten mit einer affektiven Störung diagnostiziert als in den anderen beiden Gruppen. Komorbidität war bei den untersuchten Patienten die Regel: Über 90% der Patienten wurden mit mindestens zwei Störungen (F-Diagnosen) diagnostiziert; am häufigsten (Modalwert) wurden drei F-Diagnosen gestellt. Dabei war in Gruppe 3 die durchschnittliche Anzahl an F-Diagnosen zum Aufnahmezeitpunkt signifikant geringer als in den Gruppen 1 und 2.

Keine Unterschiede zwischen den Gruppen zeigten sich bezüglich der Schwere der Beeinträchtigung bei der Aufnahme. Sowohl die patientenseitige Einschätzung der globalen psychischen Beeinträchtigung (GSI) und interpersonalen Beeinträchtigung (IIP Gesamtwert) als auch die über Therapeutenratings bewertete Beeinträchtigungsschwere (BSS) waren in den drei Gruppen vergleichbar. Zum Aufnahmezeitpunkt lag der SCL-90-Gesamtwert (GSI) bei etwa 90% der Patienten (Gesamtstichprobe) im dysfunktionalen Bereich (> 84. Perzentil der Bevölkerungsstichprobe). Auch waren keine Unterschiede zwischen den Gruppen in der Einschätzung der therapeutischen Arbeitsbeziehung (HAQ Patient) festzustellen. Auch

bezüglich der Aufenthaltsdauer, die in der Gesamtstichprobe durchschnittlich in etwa drei Monate betrug, unterschieden sich die Gruppen nicht.

Tabelle 4: Klinische Baseline-Merkmale der Patientenstichprobe

	Gruppe 1 (n=163)	Gruppe 2 (n=168)	Gruppe 3 (n=206)	Gesamt (n=537)	Gruppe 1 vs. Gruppe 2 vs. Gruppe 3
<i>Erstdiagnose*</i>					
F2, n (%)	12 (7.5)	14 (8.4)	10 (4.9)	36 (6.8)	
F3, n (%)	108 (67.9)	109 (65.7)	96 (47.3)	313 (59.3)	X <sup>2</sup> =39.887 (df=10), p=.000
F4, n (%)	33 (20.8)	28 (16.9)	68 (33.5)	129 (24.4)	
F5, n (%)	1 (0.6)	12 (7.2)	25 (12.3)	38 (7.2)	
F6, n (%)	4 (2.5)	3 (1.8)	3 (1.5)	10 (1.9)	
Sonstige, n (%)	1 (0.6)	0 (0.0)	1 (0.5)	2 (0.4)	
Komorbidität, n (%)	152 (95.6)	157 (94.6)	184 (90.6)	493 (93.4)	X <sup>2</sup> =4.110 (df=2), p=.128
Anzahl F-Diagnosen, M (sd)	3.54 (1.14)	3.44 (1.16)	3.07 (1.18)	3.33 (1.18)	F=8.320 (df=2), p=.000
<i>Beeinträchtigung</i>					
GSI, M (sd)	1.40 (0.66)	1.44 (0.66)	1.40 (0.70)	1.41 (0.68)	F=0.171 (df=2), p=.843
GSI dysfunktional (> 84. Perzentil), n (%)	147 (90.2)	158 (94.0)	184 (89.3)	489 (91.1)	X <sup>2</sup> =2.762 (df=2), p=.251
IIP Gesamt**, M (sd)	14.02 (4.17)	14.22 (3.95)	13.92 (4.55)	14.05 (4.25)	F=0.220 (df=2), p=.803
BSS Gesamt***, M (sd)	7.13 (1.78)	7.24 (1.28)	7.39 (1.95)	7.27 (1.71)	F=1.072 (df=2), p=.343
<i>Therapeutische Beziehung</i>					
HAQ Patient, M (sd)	1.46 (0.99)	1.26 (1.16)	1.44 (1.06)	1.39 (1.07)	F=1.667 (df=2), p=.190
Aufenthaltsdauer (Tage), M (sd)	96.72 (27.98)	98.83 (26.25)	96.36 (26.80)	97.24 (26.97)	F=0.431 (df=2), p=.650

*Anmerkungen:* F2 = Schizophrenie, schizotype und wahnhaftige Störungen; F3 = Affektive Störungen; F4 = Neurotische, Belastungs- und somatoforme Störungen; F5 =Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren; F6 = Persönlichkeits- und Verhaltensstörungen; GSI = Globaler Schwere-Index; IIP Gesamt = Gesamtwert Inventar Interpersonaler Probleme; BSS = Beeinträchtigungsschwere-Score; HAQ = Helping Alliance Questionnaire.

\* n=9 missings, \*\* n=5 missings, \*\*\* n=10 missings

## 3.2 Überprüfung der Fragestellungen

### 3.2.1 Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen von computergestütztem Therapeutenfeedback

#### **3.2.1.1 Praktikabilität (Umsetzbarkeit der Intervention im klinischen Routinebetrieb)**

Bei protokollkonformer Umsetzung des Untersuchungsplans hätten komplette Zwischenerhebungen von 537 Patienten sowie die dazu gehörigen Therapeutenfeedbackbögen (in den Gruppen 2 und 3 zu jedem Erhebungszeitpunkt, in Gruppe 1 nur bei der zweiten Zwischenerhebung und bei Entlassung) vorliegen müssen. Abbildung 7 sind die Anzahl der ausgefüllten Therapeutenfeedbackbögen zu entnehmen.

Bei Aufnahme fehlten zu den 374 kompletten Patientenerhebungen in den Gruppen 2 und 3 lediglich 16 Therapeuten-Feedbackbögen (wobei 4 missings durch Therapeutenwechsel bedingt waren). Bei der ersten Zwischenerhebung fehlten 22 Feedbackbögen in den Gruppen 2 und 3 (davon 2 missings aufgrund von Therapeutenwechsel). Bei protokollkonformer Umsetzung hätten bei der zweiten Zwischenerhebung 517 Patientendatensätze vorliegen müssen (20 Patienten waren zu diesem Zeitpunkt bereits entlassen worden). Bei der zweiten Zwischenerhebung waren 14 missings zu verzeichnen. Bei insgesamt 503 Patientenerhebungen in den drei Gruppen fehlten 36 Feedbackbögen. Im Therapieverlauf wurden demnach zu 1251 auf kompletten Patientendaten basierenden Rückmeldungen insgesamt 1177 Feedbackbögen von den Bezugstherapeuten ausgefüllt. Es kann davon ausgegangen werden, dass 94,1% aller Rückmeldungen von den Therapeuten betrachtet wurden. Bezüglich der Ausfüllquote der Therapeutenfeedbackbögen zeigten sich keine Unterschiede zwischen den drei Gruppen: In Gruppe 1 wurden zu 95,9% der Rückmeldungen (hier ausschließlich zur zweiten Zwischenerhebung) Feedbackbögen ausgefüllt, in Gruppe 2 waren es 94,8% und in Gruppe 3 lagen für 93,1% der Rückmeldungen Therapeuten-Feedbackbögen vor (bezogen auf die Zeitpunkte Aufnahme, erste und zweite Zwischenerhebung).

Zum Entlassungszeitpunkt wurden zu 537 Patientenerhebungen lediglich 368 Feedbackbögen ausgefüllt (68,5%). Hier unterschieden sich die Gruppen bezüglich der „Ausfüllquote“: Während in den Gruppen 1 (121 Feedbackbögen, 74,2% Ausfüllquote) und 3 (149 Feedbackbögen, 72,3% Ausfüllquote) ähnlich häufig Therapeutenbögen zu den Rückmeldungen ausgefüllt wurden, lagen für Gruppe 2 mit einer Ausfüllquote von 58,3% (98 Bögen) signifikant weniger Therapeutenfeedbackbögen vor ( $X^2=11,937$ ,  $df=2$ ,  $p=.003$ ).



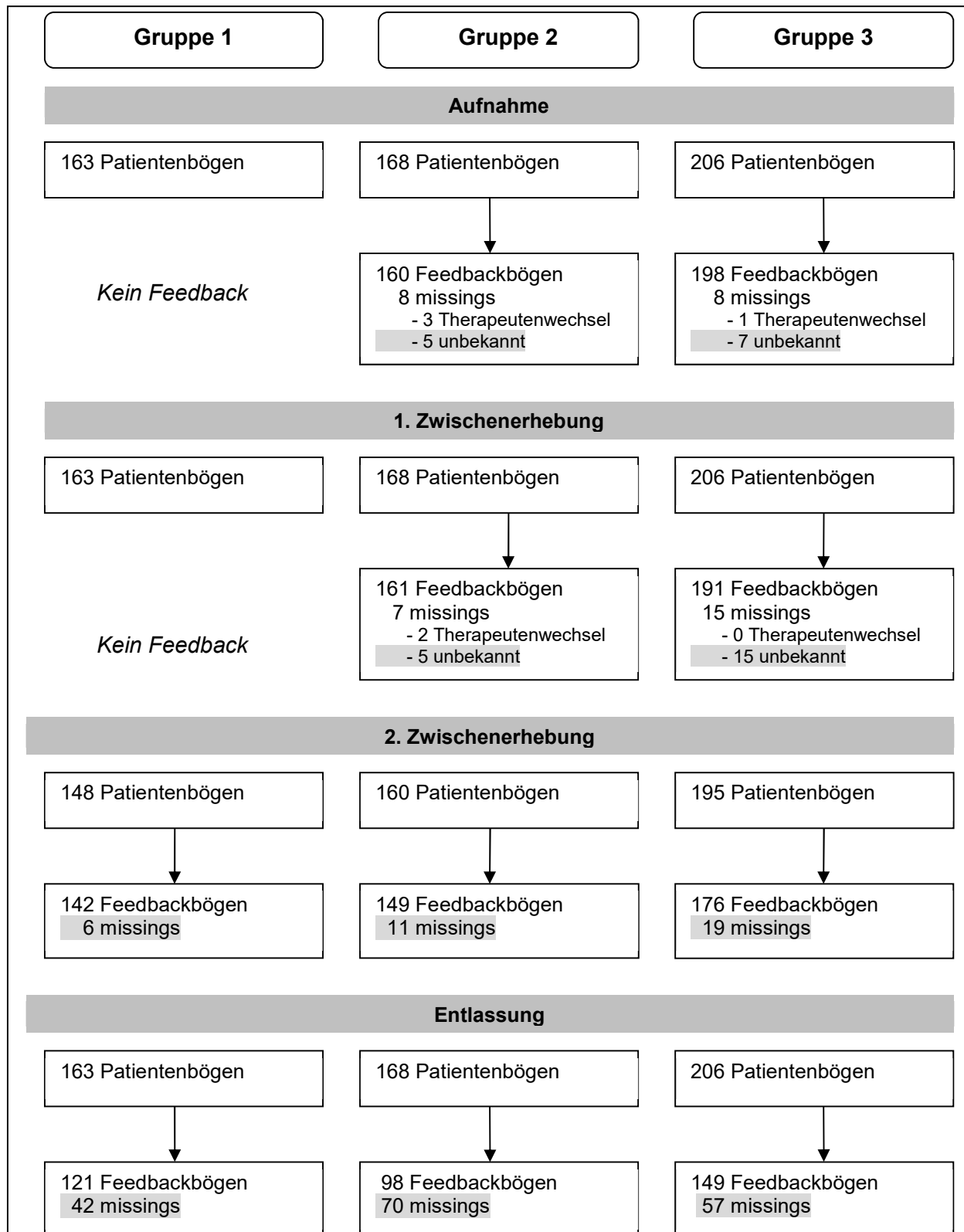


Abbildung 7: Anzahl ausgefüllter Patientenfragebögen, Therapeuten-Feedbackbögen sowie fehlender Feedbackbögen (missings) in den drei Untersuchungsgruppen zu den vier Erhebungszeitpunkten

### 3.2.1.2 Akzeptanz des Therapeutenfeedbacks

Im Feedbackbogen gaben die Therapeuten an, ob sie der Rückmeldung zustimmten (Akzeptanz), ob sie im Fall einer Zustimmung die Feedbacknachricht als Bestätigung, Ergänzung oder Veränderung ihres klinischen Eindrucks bewerteten und welche Gründe vorlagen, falls sie der Rückmeldung nicht zustimmten.

Bezüglich des Feedbacks zum Aufnahmezeitpunkt ( $X^2=2.569$ ,  $df=1$ ,  $p=.109$ ) und der ersten Zwischenerhebung zwei Wochen später ( $X^2=0.342$ ,  $df=1$ ,  $p=.559$ ) zeigte sich kein Unterschied in der „Zustimmungsrate“ zwischen den Gruppen 2 und 3.

Im Vergleich zu den beiden Gruppen mit frühem und spätem Feedback fällt die Zustimmung der Therapeuten in Gruppe 1 mit ausschließlich „spätem“ Feedback geringer aus: Sie ist sowohl zur zweiten Zwischenerhebung mit 75% als auch bei Entlassung mit 65% signifikant niedriger als in den Gruppen 2 und 3 (2. Zwischenerhebung:  $X^2=19,903$ ,  $df=2$ ,  $p=.000$ , Entlassung:  $X^2=15,259$ ,  $df=2$ ,  $p=.001$ ).

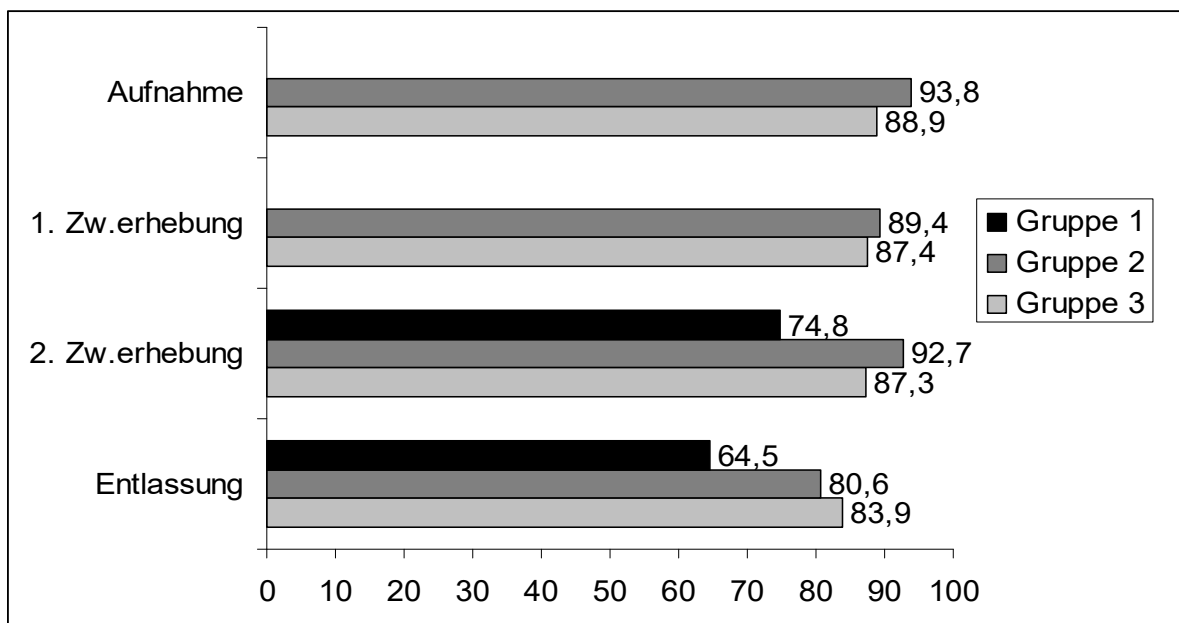


Abbildung 8: „Ich stimme der Rückmeldung zu“ – Therapeutenangaben in % für die drei Gruppen zu den vier Erhebungszeitpunkten (Anmerkung: Für Gruppe 1 lagen aufgrund des Designs – „späte“ Feedbackgruppe – lediglich Angaben zur zweiten Zwischenerhebung und bei Entlassung vor)

Tabelle 5 zeigt, dass die Rückmeldungen am häufigsten als Bestätigung des klinischen Eindrucks gewertet wurden, in 40% aller Fälle (531 von insgesamt 1327 Feedbackbögen mit Zustimmung) auch als das klinische Urteil ergänzende Information. Nur sehr selten wurde angegeben, dass der klinische Eindruck durch die Rückmeldungen verändert worden sei.

Falls die Therapeuten der Rückmeldung nicht zustimmten, konnten als mögliche Gründe angegeben werden, dass die Fragebögen die Beeinträchtigung des Patienten unterschätzen oder überschätzen würden, dass die Fragebögen bestimmte Probleme des Patienten nicht berücksichtigt hätten oder sonstige Gründe für die Ablehnung der Rückmeldung vorlägen.

**Tabelle 5:** Zustimmung zur Rückmeldung: Bestätigung, Ergänzung oder Veränderung des klinischen Eindrucks der Bezugstherapeuten zu den vier Erhebungszeitpunkten (Mehrfachantworten möglich).

	Gruppe 1	Gruppe 2	Gruppe 3
<i>Aufnahme</i>		<i>N=150</i>	<i>N=176</i>
Bestätigung, n (%)		133 (88.7)	126 (71.6)
Ergänzung, n (%)	<i>kein Feedback</i>	46 (30.7)	80 (45.5)
Veränderung, n (%)		4 (2.7)	5 (2.8)
<i>1. Zw.erhebung</i>		<i>N=144</i>	<i>N=167</i>
Bestätigung, n (%)		126 (87.5)	105 (62.9)
Ergänzung, n (%)	<i>kein Feedback</i>	44 (30.6)	81 (48.5)
Veränderung, n (%)		5 (3.5)	6 (3.6)
<i>2. Zw.erhebung</i>	<i>N=110</i>	<i>N=140</i>	<i>N=158</i>
Bestätigung, n (%)	74 (67.3)	128 (91.4)	112 (70.9)
Ergänzung, n (%)	55 (50.0)	41 (29.3)	72 (45.6)
Veränderung, n (%)	6 (5.5)	2 (1.4)	7 (4.4)
<i>Entlassung</i>	<i>N=78</i>	<i>N=79</i>	<i>N=125</i>
Bestätigung, n (%)	59 (75.6)	73 (92.4)	92 (73.6)
Ergänzung, n (%)	39 (50.0)	22 (27.8)	51 (40.8)
Veränderung, n (%)	2 (2.6)	1 (1.3)	2 (1.6)

Tabelle 6 zeigt, dass in den Gruppen 2 und 3 als häufigster Grund genannt wurde, dass die psychometrischen Instrumente die Beeinträchtigung des Patienten unterschätzen würden, d.h. der Therapeut die Beeinträchtigung des Patienten als gravierender einschätzte als der Patient selbst. Dagegen wurde in Gruppe 1 als häufigster Grund genannt, die Fragebögen würden Probleme des Patienten nicht berücksichtigen. In 19 Fällen wurden sonstige Gründe für die Ablehnung der Feedbacknachricht angegeben. Dabei wurden je vier Mal Versorgungswünsche bzw. ein Rentenbegehren des Patienten als Motivation für die Beantwortung genannt sowie eine Dissimulierung der Problematik durch den Patienten.

**Tabelle 6:** Keine Zustimmung zur Rückmeldung: Gründe für die Ablehnung des Feedbacks (Fb) (Mehrfachantworten möglich)

	Gruppe 1	Gruppe 2	Gruppe 3
<i>N (Aufnahme, Zwischenerhebungen und Entlassung zusammengefasst)</i>	<i>N=80</i>	<i>N=57</i>	<i>N=93</i>
Fb unterschätzt Beeinträchtigung, n (%)	28 (35.0)	42 (73.7)	69 (74.2)
Fb überschätzt Beeinträchtigung, n (%)	25 (31.3)	10 (17.5)	24 (25.8)
Fb bildet zentrale Probleme nicht ab, n (%)	36 (45.0)	17 (29.8)	8 (8.6)
Sonstiges	8 (10.0)	9 (15.8)	2 (2.2)

### 3.2.1.3 Relevanz und Nutzen des Therapeutenfeedbacks

Bei Aufnahme und zu den Zwischenerhebungen wurden die Therapeuten um ihre Einschätzung gebeten, in wie weit die Rückmeldungen relevante Informationen enthalten. Dabei waren die Antwortformate „überhaupt nicht“, „etwas“, „ziemlich“ und „sehr“ vorgegeben. Für Gruppe 1 lagen aufgrund des Studiendesigns nur Angaben zur zweiten Zwischenerhebung vor. Am häufigsten wurden die in den Rückmeldungen enthaltenen Informationen als „etwas relevant“ eingeschätzt (66,8% aller Feedbacknachrichten) und nur äußerst selten – in insgesamt 0,5% aller Fälle - als „sehr relevant“.

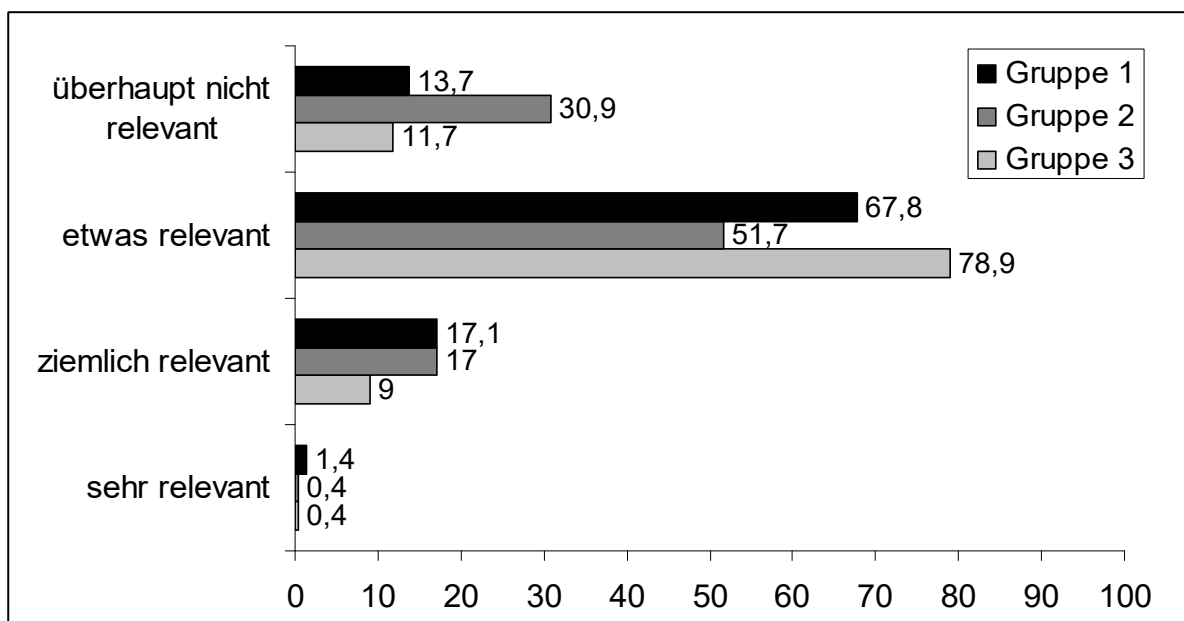


Abbildung 9: „Enthält die Rückmeldung relevante Informationen?“ Therapeutenangaben in Prozent, Zeitpunkte Aufnahme, erste und zweite Zwischenerhebung zusammengefasst

Bei Aufnahme und bei der ersten Zwischenerhebung wurden die Feedbackinformationen von Therapeuten der Gruppe 2 signifikant häufiger als „überhaupt nicht relevant“ und signifikant seltener als „etwas relevant“ bewertet als von Therapeuten der Gruppe 3 (Aufnahme:  $X^2=35.094$ ,  $df=3$ ,  $p=.000$ ; 1. Zwischenerhebung:  $X^2=30.385$ ,  $df=3$ ,  $p=.000$ ). Auch bei der zweiten Zwischenerhebung zeigten sich signifikante Unterschiede in den Bewertungen – dieses Mal hinsichtlich der zu diesem Erhebungszeitpunkt befragten drei Untersuchungsgruppen ( $X^2=31.206$ ,  $df=6$ ,  $p=.000$ ), wobei wiederum in Gruppe 2 die Einschätzung „überhaupt nicht relevant“ häufiger und „etwas relevant“ seltener abgegeben wurden im Vergleich zu Therapeuten der Gruppen 1 und 3 .

Bei Aufnahme und zu den Zwischenerhebungen wurde außerdem erfragt, in wie weit die Therapeuten die Rückmeldung als hilfreich für die weitere Planung des Therapieverlaufs einschätzten. Es waren wiederum die Antwortformate „überhaupt nicht“, „etwas“, „ziemlich“ und „sehr“ vorgegeben. Die über die drei Erhebungszeitpunkte hinweg zusammengefassten

Antworten sind Abbildung 10 zu entnehmen (Für die Gruppe 1 gingen wiederum nur die Daten der zweiten Zwischenerhebung ein).

In allen Feedbackgruppen wurden die Rückmeldungen am häufigsten als „etwas hilfreich“ und nur äußerst selten – in weniger als 1% der Fälle – als „sehr hilfreich“ für die Planung des weiteren Behandlungsverlaufs bewertet.

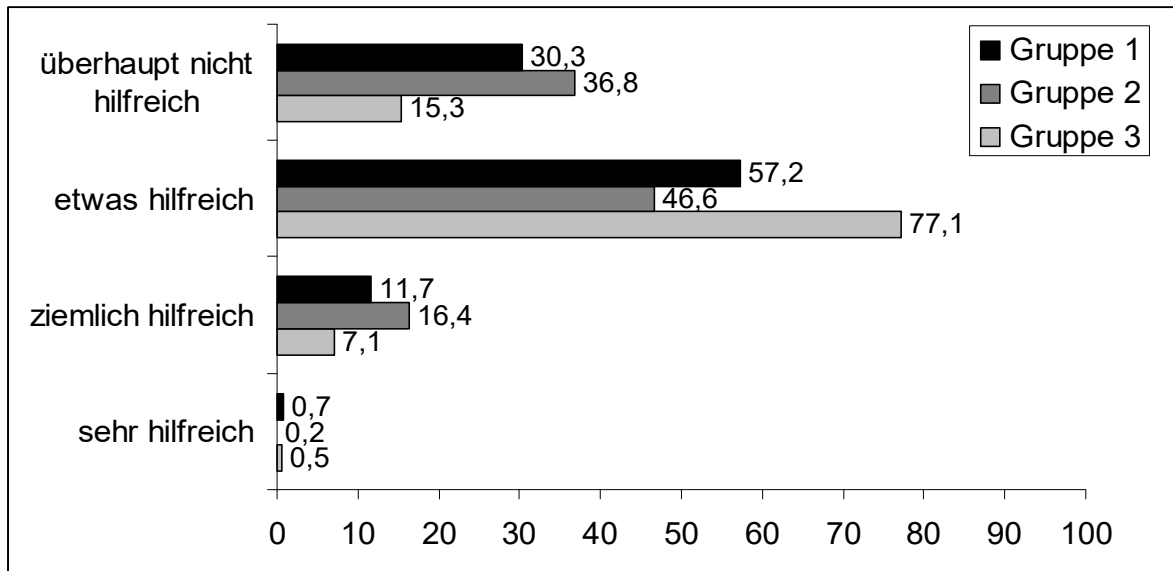


Abbildung 10: „Ist die Rückmeldung hilfreich für die weitere Behandlung?“ Therapeutenangaben in Prozent, Zeitpunkte Aufnahme, erste und zweite Zwischenerhebung zusammengefasst

Therapeuten in Gruppe 3 bewerteten im Vergleich zu Therapeuten in den Gruppen 1 und 2 die Rückmeldungen signifikant seltener als „überhaupt nicht hilfreich“ und häufiger als „etwas hilfreich“. Dieser signifikante Unterschied zeigte sich zu jedem der drei Erhebungszeitpunkte: Bei Aufnahme ( $X^2=43.147$ ,  $df=3$ ,  $p=.000$ ) und der ersten Zwischenerhebung ( $X^2=33.123$ ,  $df=3$ ,  $p=.000$ ) zwischen den Gruppen 2 und 3 sowie bei der zweiten Zwischenerhebung zwischen den drei Untersuchungsgruppen ( $X^2=34.329$ ,  $df=6$ ,  $p=.000$ ).

### 3.2.2 Fragestellung 2: Vergleich des Therapie-Outcomes

Unterschiede zwischen den Gruppen wurden bzgl. der psychischen Beeinträchtigung (globaler Schwereindex GSI der SCL-90 bei Entlassung), des Prozentsatzes an reliablen Verbesserungen (bezogen auf den GSI) und des Auffälligkeitssignals als aggregiertes Outcome-Maß getestet. Hypothesengemäß wurde erwartet, dass Patienten, die bei der ersten Zwischenerhebung als „auffällig“ klassifiziert worden waren (vgl. Abbildung 5, S. X), in den Gruppen 2 und 3 - in denen zu diesem Zeitpunkt Verlaufsfeedback gegeben wurde - ein besseres Therapie-Outcome erzielen als bei der ersten Zwischenerhebung auffällige Patienten in Gruppe 1. Insgesamt wurden 22,7% der Patienten ( $n=122$  von 537 Fällen) bei der ersten Zwischenerhebung als „auffällig“ klassifiziert. Dabei zeigten sich keine Unterschiede zwischen den Gruppen: In Gruppe 1 wiesen 25,8% ( $n=42$ ) der Patienten einen frühen

auffälligen Verlauf auf, in Gruppe 2 waren dies 22,6% (n=38) und in Gruppe 3 20,4% (n=42) der Patienten ( $X^2=1.501$ ,  $df=2$ ,  $p=0.472$ )

### 3.2.2.1 Test auf Gruppenunterschiede im Globalen Schwereindex GSI bei Entlassung

Tabelle 7 zeigt die Mittelwerte und Standardabweichungen des Gesamtwerts der SCL-90-R (GSI) bei Aufnahme und Entlassung für die drei Gruppen sowie die Gesamtstichprobe. Außerdem werden die GSI-Werte der drei Gruppen, unterteilt nach auffälligem oder unauffälligem Verlauf bei der ersten Zwischenerhebung dargestellt.

Tabelle 7: Mittelwerte (M), Standardabweichungen (sd) und Effektgrößen (ES) des Globalen Schwere-Index (GSI) für die Gesamtstichprobe, die drei Untersuchungsgruppen und nach Auffälligkeitsstatus bei der ersten Zwischenerhebung

		GSI Aufnahme	GSI Entlassung	ES
		M (sd)	M (sd)	
<b>Gesamt (n=537)</b>		1.41 (0.68)	0.92 (0.67)	0.72
Gruppe 1 (n=163)		1.40 (0.66)	0.96 (0.74)	0.67
Gruppe 2 (n=168)		1.44 (0.66)	0.88 (0.62)	0.85
Gruppe 3 (n=206)		1.40 (0.70)	0.92 (0.66)	0.69
<b>früher auffälliger Verlauf</b>	Gruppe 1 (n=42)	1.48 (0.71)	1.36 (0.92)	0.17
	Gruppe 2 (n=38)	1.46 (0.63)	1.03 (0.65)	0.68
	Gruppe 3 (n=42)	1.64 (0.81)	1.24 (0.69)	0.49
<b>früher unauffälliger Verlauf</b>	Gruppe 1 (n=121)	1.38 (0.65)	0.82 (0.61)	0.86
	Gruppe 2 (n=130)	1.43 (0.67)	0.84 (0.61)	0.88
	Gruppe 3 (n=164)	1.34 (0.67)	0.84 (0.63)	0.75

Gruppenunterschiede bezüglich der abhängigen Variable „GSI bei Entlassung“ wurden mittels gemischter linearer Modelle geprüft. Hypothesenkonform zeigte sich für die Interaktion zwischen Gruppe und frühem auffälligen Verlauf ein signifikanter Effekt (Tabelle 8). Auch für die Kovariaten „GSI bei Aufnahme“ und „früher auffälliger Verlauf“ (Auffälligkeitssignal bei erster Zwischenerhebung) ergaben sich signifikante Effekte.

Tabelle 8: Gemischtes lineares Modell für abhängige Variable „GSI bei Entlassung“

	F-Wert (df1; df2)	p-Wert
Gruppe	1.87 (2; 21.9)	0.1774
früher auffälliger Verlauf	28.43 (1; 527)	< 0.0001
Gruppe*früher auffälliger Verlauf	3.33 (2; 527)	0.0367
Aufenthaltsdauer	3.67 (1; 529)	0.0559
GSI Aufnahme	323.50 (1; 526)	< 0.0001

Es wurden Einzelvergleiche mit Bonferroni-Korrektur durchgeführt, um zu untersuchen, welche der drei Gruppen sich bezüglich des Kriteriums „GSI bei Entlassung“ unterscheiden. Es zeigte sich ein signifikanter Unterschied zwischen den GSI-Werten bei den jeweils zur ersten Zwischenerhebung auffälligen Patienten der Gruppen 1 und 2 (geschätzte Differenz -0.335,  $p=0.01$ ). Entgegen der Hypothese ergaben sich keine signifikanten Unterschiede im GSI bei den jeweils zur ersten Zwischenerhebung auffälligen Patienten der Gruppen 1 und 3 (geschätzte Differenz -0.198,  $p=0.12$ ). Wie erwartet unterschieden sich die GSI-Werte bei Entlassung nicht signifikant bei „früh auffälligen“ Patienten der Gruppen 2 und 3 (geschätzte Differenz 0.137,  $p=0.29$ ).

### 3.2.2.2 Test auf Gruppenunterschiede bezüglich reliabler Veränderungen im GSI

Die Raten an reliablen Verbesserungen und Verschlechterungen auf den neun Subskalen der SCL-90-R für die drei Gruppen sowie die Gesamtstichprobe sind in Anhang C aufgeführt. Das höchste Ausmaß an reliablen Veränderungen zeigt sich auf dem Gesamtwert GSI: In der Gesamtstichprobe verbesserten sich die GSI-Werte von 70% der Patienten, während sich bei 9,5% der Patienten die Werte zwischen Aufnahme und Entlassung verschlechterten. Auf den Subskalen zeigt sich die höchste Verbesserungsrate bei „Depressivität“ (SCL 4); hier verbesserten sich die Werte zwischen Aufnahme und Entlassung bei fast zwei Drittel der Patienten (62,4%). Das geringste Ausmaß an Veränderungen ist auf den Skalen „Somatisierung“ (SCL 1) und „Aggressivität/Feindseligkeit“ (SCL 6) zu beobachten; hier zeigen sich reliable Verbesserungen nur bei etwa einem Drittel der Patienten.

Abbildung 11 zeigt den Prozentsatz an Patienten mit reliabler Verbesserung, reliabler Verschlechterung oder keiner Veränderung bezüglich des Globalen Schwereindex GSI von der Aufnahme bis zur Entlassung, jeweils für Patienten mit frühem auffälligem und unauffälligem Behandlungsverlauf. In allen drei Untersuchungsgruppen zeigte sich bei etwa drei Viertel der Patienten, die zwei Wochen nach Aufnahme als „unauffällig“ klassifiziert worden waren, eine reliable Verbesserung der psychischen Symptomatik bis zur Entlassung.

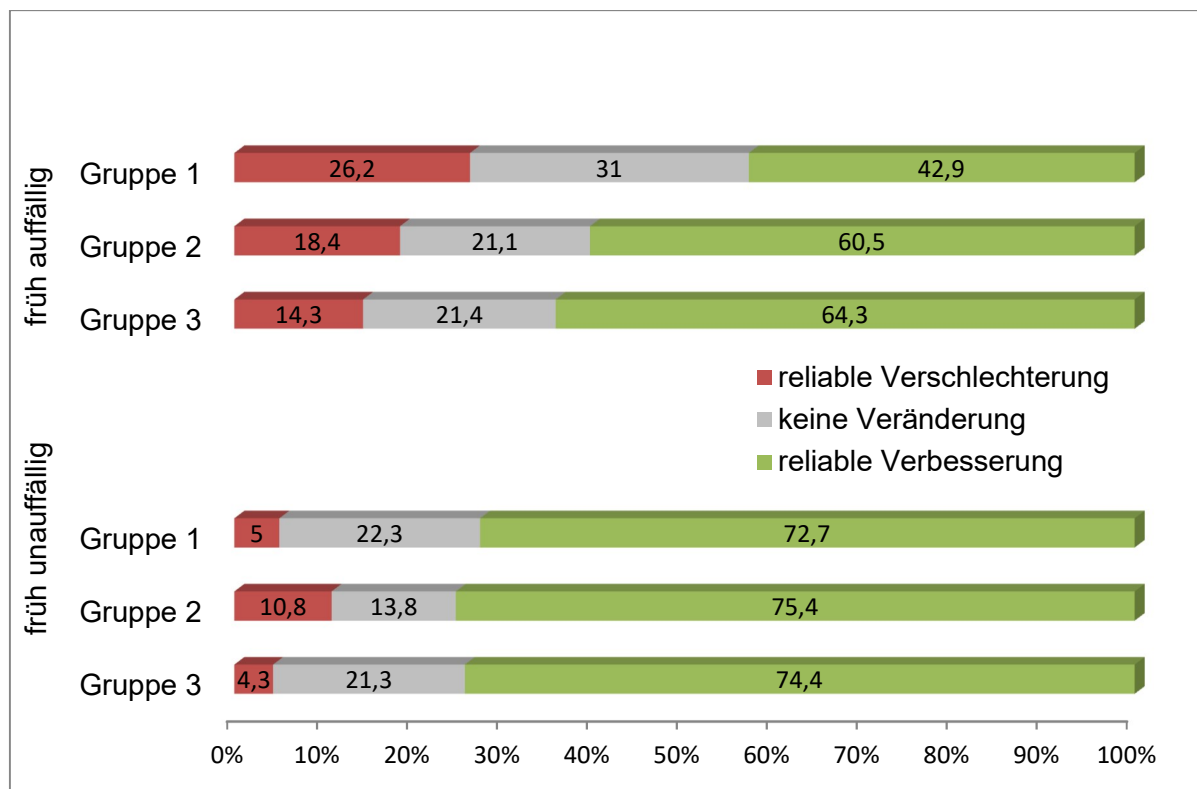


Abbildung 11: Reliable Veränderungen des Globalen Schwereindex (GSI) von Aufnahme bis Entlassung in Prozent, jeweils für Patienten mit und ohne Auffälligkeitssignal bei der ersten Zwischenerhebung

Bei Patienten mit einem frühen auffälligen Verlauf erzielten jeweils 61% (Gruppe 2) bzw. 64% (Gruppe 3) der Patienten, deren Therapeuten“ zu diesem Zeitpunkt Feedback erhalten hatten, bis zur Entlassung eine reliable Verbesserung ihres Gesundheitszustands. Bei Patienten der Gruppe 1 (Verlaufsfeedback an die Therapeuten 6 Wochen nach Aufnahme) waren dies 43% der Patienten mit einem frühen auffälligen Verlauf.

Gruppenunterschiede bezüglich der abhängigen Variable „Reliable Veränderung des GSI“ wurden durch ein gemischtes logistisches Regressionsmodell geschätzt (Tabelle 9). Dabei wurden die Bedingungen „keine reliable Veränderung“ und „reliable Verschlechterung“ zusammengefasst und auf „0“ gesetzt, reliable Verbesserung wurde als „1“ kodiert.

Tabelle 9: Gemischtes logistisches Regressionsmodell mit abhängiger Variable „reliable Veränderung des GSI“

	F-Wert (df1; df2)	p-Wert
Gruppe	1.18 (2; 22.08)	0.325
früher auffälliger Verlauf	17.77 (1; 529)	< 0.0001
Gruppe*früher auffälliger Verlauf	1.01 (2; 529)	0.365
Aufenthaltsdauer	3.92 (1; 529)	0.048
GSI Aufnahme	20.64 (1; 529)	< 0.0001



Entgegen der Hypothese zeigte sich kein Interaktionseffekt zwischen den Variablen „Gruppe“ und „früher auffälliger Verlauf“. Ebenso zeigte sich kein Haupteffekt für „Gruppe“. Signifikante Effekte ergaben sich für die Kovariaten „GSI bei Aufnahme“, „Aufenthaltsdauer“ und „früher auffälliger Verlauf“.

### 3.2.2.3 Test auf Gruppenunterschiede im Auffälligkeitssignal

Insgesamt ergab der in Abbildung 5 (S. 47) dargestellte Algorithmus für 25,9% (n=139) aller Patienten ein Auffälligkeitssignal bei Entlassung. Es wurden 29,4% der Patienten in Gruppe 1, 25% der Patienten in Gruppe 2 und 23,8% der Patienten in Gruppe 3 als „auffällig“ klassifiziert. In Abbildung 12 ist der Prozentsatz an auffälligen und unauffälligen (positiven) Verläufen zum Entlassungszeitpunkt, jeweils für Patienten mit und ohne Auffälligkeitssignal bei der ersten Zwischenerhebung, dargestellt. In allen drei Gruppen erzielten knapp 80 Prozent der Patienten mit einem frühen unauffälligen Verlauf auch einen insgesamt positiven (unauffälligen) Behandlungsverlauf. Heterogener ist das Ergebnis bei Patienten, die bei der ersten Zwischenerhebung als auffällig klassifiziert wurden.

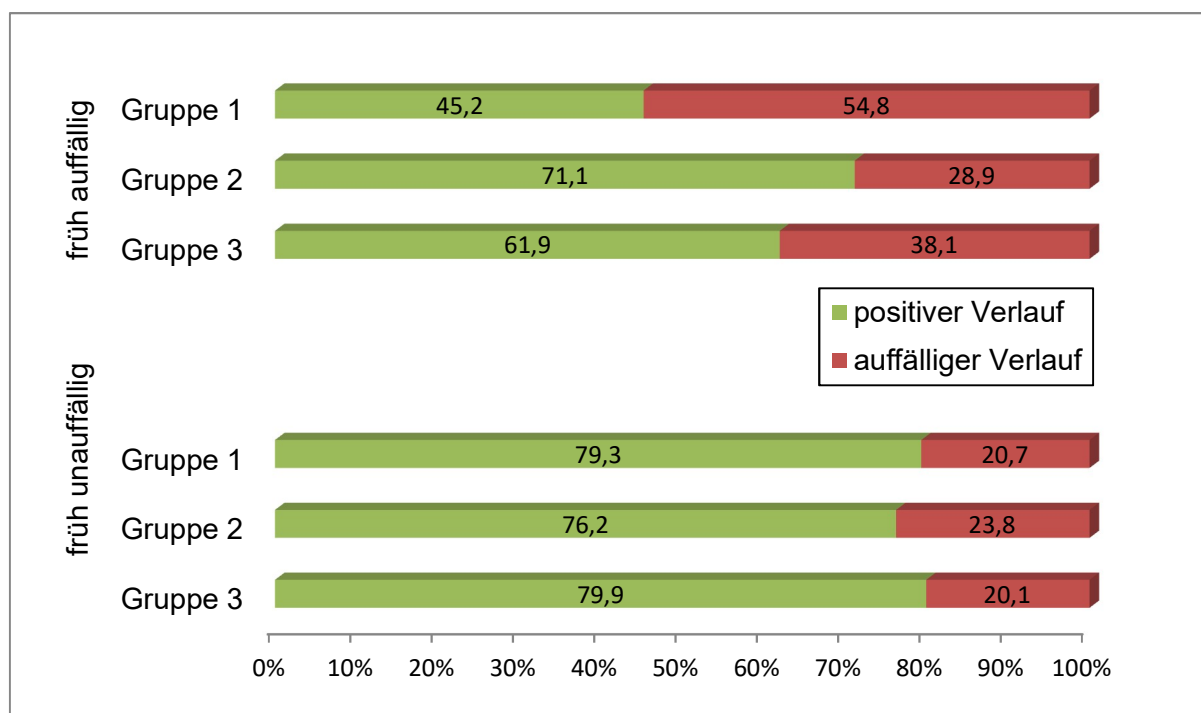


Abbildung 12: Auffällige und unauffällige (positive) Verläufe bei Entlassung (Prozentangaben), jeweils für Patienten mit und ohne Auffälligkeitssignal bei der ersten Zwischenerhebung

Gruppenunterschiede bezüglich der abhängigen Variable „Auffälligkeitssignal“ wurden ebenfalls mittels eines gemischten logistischen Regressionsmodells geprüft (Tabelle 9). Hypothesenkonform zeigte sich ein signifikanter Effekt für die Interaktion „Gruppe \* früher auffälliger Verlauf“. Signifikante Effekte ergaben sich für die Kovariaten „früher auffälliger Verlauf“ und „Aufenthaltsdauer“.

*Tabelle 10:* Gemischte logistisches Regressionsmodell mit abhängiger Variable „Auffälligkeitssignal bei Entlassung“

	F-Wert (df1; df2)	p-Wert
Gruppe	0.94 (2; 21.42)	0.405
früher auffälliger Verlauf	15.50 (1; 529)	< 0.000
Gruppe*früher auffälliger Verlauf	3.36 (2; 529)	0.035
Aufenthaltsdauer	5.09 (1; 529)	0.025
GSI Aufnahme	1.17 (1; 529)	0.281

Es wurden Einzelvergleiche mit Bonferroni-Holm-Adjustierung durchgeführt, um zu untersuchen, welche der drei Gruppen sich bezüglich der abhängigen Variable unterscheiden. Es zeigte sich ein signifikanter Unterschied zwischen den jeweils zur ersten Zwischenerhebung auffälligen Patienten der Gruppen 1 und 2 (-1.344,  $p=0.03$ ,  $OR=0.261$ ,  $KI=0.080-0.851$ ). Entgegen der Hypothese ergaben sich keine signifikanten Unterschiede bezüglich des Auffälligkeitssignals bei Entlassung bei den zur ersten Zwischenerhebung auffälligen Patienten der Gruppen 1 und 3 (-0.647,  $p=0.24$ ,  $OR=0.524$ ,  $KI=0.174-1.571$ ). Wie erwartet unterschieden sich die Patienten mit einem frühen auffälligen Verlauf der Gruppen 2 und 3 nicht bezüglich des Auffälligkeitssignals bei Entlassung (0.697,  $p=0.24$ ,  $OR=2.007$ ,  $KI=0.616-6.542$ ).

### 3.2.3 Fragestellung 3: Vergleich der Behandlungsdauer

In der Gesamtstichprobe betrug die durchschnittliche Behandlungsdauer 97,24 Tage ( $sd=26,97$ ). In den drei Untersuchungsgruppen wurden die Patienten durchschnittlich 96,72 Tage ( $sd=27,98$ , Gruppe 1), 98,83 Tage ( $sd=26,25$ , Gruppe 2) und 96,36 Tage ( $sd=26,80$ , Gruppe 3) behandelt.

#### **3.2.3.1 Test auf Gruppenunterschiede bezüglich der Behandlungsdauer bei Patienten mit frühem auffälligem Verlauf**

Es wurde erwartet, dass Patienten, die zwei Wochen nach Aufnahme als „auffällig“ klassifiziert wurden, in den Gruppen 2 und 3 länger behandelt werden als Patienten in Gruppe 1 mit einem Auffälligkeitssignal bei der ersten Zwischenerhebung.

In Tabelle 11 sind die Behandlungsdauern von Patienten der drei Untersuchungsgruppen dargestellt, getrennt nach dem Signal „auffälliger Verlauf“ versus „unauffälliger Verlauf“ bei der ersten Zwischenerhebung.

*Tabelle 11:* Aufenthaltsdauer in den drei Untersuchungsgruppen in Abhängigkeit von einem frühen auffälligen Verlauf oder keinem Auffälligkeitssignal bei der ersten Zwischenerhebung.

<b>Auffälligkeitssignal bei 1. Zwischenerhebung</b>		<b>Aufenthaltsdauer</b>	
		Md	M (sd)
<b>auffällig</b>	<i>Gruppe 1</i>	96.00	97.95 (28.73)
	<i>Gruppe 2</i>	105.00	104.13 (28.85)
	<i>Gruppe 3</i>	105.00	108.00 (30.29)
<b>Nicht auffällig</b>	<i>Gruppe 1</i>	100.00	96.29 (27.82)
	<i>Gruppe 2</i>	98.5	97.28 (25.35)
	<i>Gruppe 3</i>	96.00	93.38 (25.07)

*Anmerkungen:* Md: Median, M: Mittelwert, sd: Standardabweichung.

Es wurde ein gemischtes lineares Modell mit „Behandlungsdauer“ als abhängiger Variable und den Prädiktoren „Gruppe“ und „früher auffälliger Verlauf“ geschätzt; entsprechend der Hypothese wurde die Interaktion „Gruppe \* früher auffälliger Verlauf“ in das Modell aufgenommen, Als Kovariate wurde „GSI bei Aufnahme“ berücksichtigt. (Tabelle 12).

*Tabelle 12:* Gemischtes lineares Modell mit abhängiger Variable „Behandlungsdauer“

	F-Wert (df1; df2)	p-Wert
Gruppe	0.49 (2; 16.9)	0.6235
früher auffälliger Verlauf	5.83 (1; 526)	0.0161
Gruppe*früher auffälliger Verlauf	1.53 (2; 526)	0.2175
GSI Aufnahme	5.64 (1; 525)	0.0179

Entgegen der Hypothese zeigte sich kein Interaktionseffekt, d.h. Patienten mit einem frühen auffälligen Verlauf, deren Therapeuten darüber Rückmeldung erhalten hatten (Gruppen 2 und 3), wurden nicht länger behandelt als Patienten der Gruppe 1 mit einem frühen auffälligen Verlauf. Auch zeigte sich kein Haupteffekt für „Gruppe“.

Signifikante Effekte zeigten sich für die Kovariaten „früher auffälliger Verlauf“ und „GSI bei Aufnahme“: Patienten mit frühem auffälligen Verlauf wurden mit 102,44 Tagen (geschätzter Randmittelwert, KI: 96,3 – 107,95) länger behandelt als Patienten mit unauffälligem frühen Behandlungsverlauf (geschätzter Randmittelwert 95,68, KI: 91,61 – 99,74).

### 3.2.3.2 Test auf Gruppenunterschiede bezüglich der Behandlungsdauer bei Patienten, deren GSI bei der ersten Zwischenerhebung im Normbereich lag

Es wurde angenommen, dass Patienten, deren psychische Gesamtbelastung zwei Wochen nach Aufnahme innerhalb des Normbereichs („funktionaler“ Bereich) liegt, in den Gruppen 2 und 3 kürzer behandelt werden als Patienten in Gruppe 1.

*Tabelle 13:* Aufenthaltsdauer (in Tagen) für Patienten, deren GSI bei der ersten Zwischenerhebung innerhalb („funktional“) versus außerhalb („dysfunktional“) des Normbereichs lag.

GSI bei 1. Zwischenerhebung		Aufenthaltsdauer	
		Md	M (sd)
im „funktionalen“ Bereich	Gruppe 1	92.00	85.58 (30.69)
	Gruppe 2	87.50	86.58 (21.40)
	Gruppe 3	89.00	86.29 (22.67)
im „dysfunktionalen“ Bereich	Gruppe 1	100.00	99.34 (26.87)
	Gruppe 2	100.00	101.07 (26.50)
	Gruppe 3	98.00	99.17 (27.25)

*Anmerkungen:* Md: Median, M: Mittelwert, sd: Standardabweichung.

Unterschiede bezüglich der abhängigen Variable „Behandlungsdauer“ wurden wiederum mittels eines gemischten linearen Modells geschätzt.

Es ergab sich weder ein signifikanter Haupteffekt für „Gruppe“ noch - entgegen der Hypothese - ein signifikanter Effekt der Interaktion „Gruppe \* GSI im Normbereich bei 1. Zwischenerhebung“. Der einzige signifikante Effekt zeigte sich bei der Kovariate „psychische Belastung im Normbereich bei 1. Zwischenerhebung“ (Tabelle 14): Patienten, deren GSI bei der 1. Zwischenerhebung im „gesunden“ Normbereich lag, wurden mit 86,74 Tagen (geschätzter Randmittelwert, KI: 80,11 – 93,34) kürzer behandelt als Patienten, deren GSI bei der ersten Zwischenerhebung im krankheitswertigen Bereich lag (geschätzter Randmittelwert 99,67, KI: 95,58 – 103,77).

*Tabelle 14:* Gemischtes lineares Modell mit abhängiger Variable „Behandlungsdauer“

	F-Wert (df1; df2)	p-Wert
Gruppe	0.14 (2; 20.5)	0.8685
GSI Normbereich 1. Zwischenerhebung	14.09 (1; 521)	0.0002
Gruppe* GSI Normbereich 1. Zwischenerhebung	0.09 (2; 522)	0.9154
GSI Aufnahme	0.07 (1; 524)	0.7971

### 3.2.4 Fragestellung 4: Vergleich der therapeutischen Beziehung

Die Mittelwerte der Beziehungsskala des HAQ sind über die vier Erhebungszeitpunkte hinweg recht konstant (Tabelle 15).

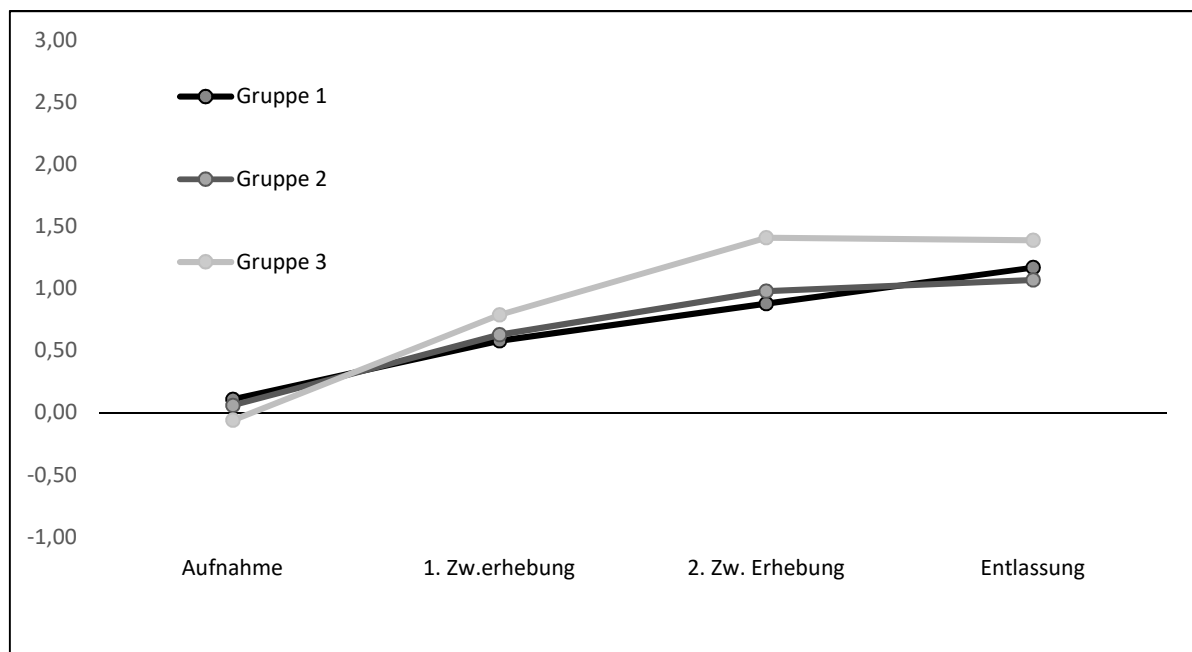
*Tabelle 15:* Mittelwerte (M) und Standardabweichungen (sd) der HAQ-Beziehungswerte für die Gesamtstichprobe und die drei Untersuchungsgruppen

	HAQ Aufnahme M (sd)	HAQ ZP 1 M (sd)	HAQ ZP2 M (sd)	HAQ Entlassung M (sd)
Gesamt (n=537)	1.39 (1.07)	1.66 (0.92)	1.74 (1.02)	1.65 (1.30)
Gruppe 1 (n=163)	1.47 (0.92)	1.74 (0.91)	1.80 (0,98)	1.72 (1.29)
Gruppe 2 (n=168)	1.29 (1.15)	1.54 (0.97)	1.69 (1,09)	1.50 (1.41)
Gruppe 3 (n=206)	1.41 (1.05)	1.69 (0.89)	1.72 (1.03)	1.89 (1.06)

Anmerkung: ZP: Zwischenerhebung

#### 3.2.4.1 Test auf Gruppenunterschiede bezüglich des Entlassungswerts der Beziehungsskala

Abbildung 13 zeigt die Mittelwerte der HAQ-Beziehungsskala über den Behandlungsverlauf für Patienten mit unterdurchschnittlichen Beziehungswerten (HAQ-Beziehungsskala < 16. Perzentil) bei Aufnahme und/oder der ersten Zwischenerhebung.



*Abbildung 13:* Mittelwerte der HAQ-Beziehungsskala zu den vier Erhebungszeitpunkten für Patienten mit initial unterdurchschnittlichen Beziehungswerten

Spezifische Effekte bezüglich der therapeutischen Beziehung wurden für die Subgruppe von Patienten angenommen, die die Beziehung zu ihrem Therapeuten im frühen Therapieverlauf als unterdurchschnittlich bewertet hatten. Entsprechend klassifiziert wurden Patienten, deren HAQ-Beziehungswert bei der Aufnahme und/oder der ersten Zwischenerhebung unter dem 16. Perzentil der klinikinternen Vergleichsstichprobe lag. Dies war bei insgesamt n=115 Patienten der Fall (n=28 in Gruppe 1, n=47 in Gruppe 2, n=40 in Gruppe 3).

Unterschiede zwischen den Gruppen bezüglich der abhängigen Variable „Beziehungsskala HAQ bei Entlassung“ wurden mittels eines gemischten linearen Modells geprüft (Tabelle 16).

*Tabelle 16:* Gemischtes lineares Modell mit abhängiger Variable „Beziehungsskala HAQ bei Entlassung“

	F-Wert (df1; df2)	p-Wert
Gruppe	1.20 (2; 17.61)	0.324
Th. Beziehung < 16. Perzentil	22.56 (1; 509.66)	0.000
Gruppe* Th. Beziehung < 16. Perzentil	1.29 (2; 509.28)	0.275
Aufenthaltsdauer	5.33 (1; 514.47)	0.021
GSI Aufnahme	0.00 (1; 506.83)	0.990

Entgegen der Hypothese zeigte sich kein signifikanter Interaktionseffekt zwischen den Variablen „Gruppe“ und „initial unterdurchschnittliche therapeutische Beziehung (Wert der HAQ-Beziehungsskala bei Aufnahme und/oder erster Zwischenerhebung <16. Perzentil)“. Außerdem zeigten sich weder ein signifikanter Haupteffekt für „Gruppe“ noch ein signifikanter Effekt für die Kovariate „GSI bei Aufnahme“. Signifikante Effekte zeigten sich für die Variablen „initial unterdurchschnittliche therapeutische Beziehung (bei Aufnahme und/oder erster Zwischenerhebung <16. Perzentil)“ sowie für die Kovariate „Aufenthaltsdauer“.

#### **3.2.4.2 Test auf Gruppenunterschiede bezüglich reliabler Veränderung der therapeutischen Beziehung**

In Abbildung 14 sind die Prozentwerte an reliablen Veränderungen (Verbesserungen, Verschlechterungen, keine Veränderungen) bezüglich der Beziehungsskala des HAQ zwischen Aufnahme und Entlassung dargestellt, jeweils für Patienten mit initial unterdurchschnittlichen Beziehungswerten (Werte der HAQ-Beziehungsskala bei Aufnahme und/oder erster Zwischenerhebung <16. Perzentil) und (über-)durchschnittlichen initialen Beziehungswerten (Werte der HAQ-Beziehungsskala bei Aufnahme und/oder erster Zwischenerhebung  $\geq$  16. Perzentil).

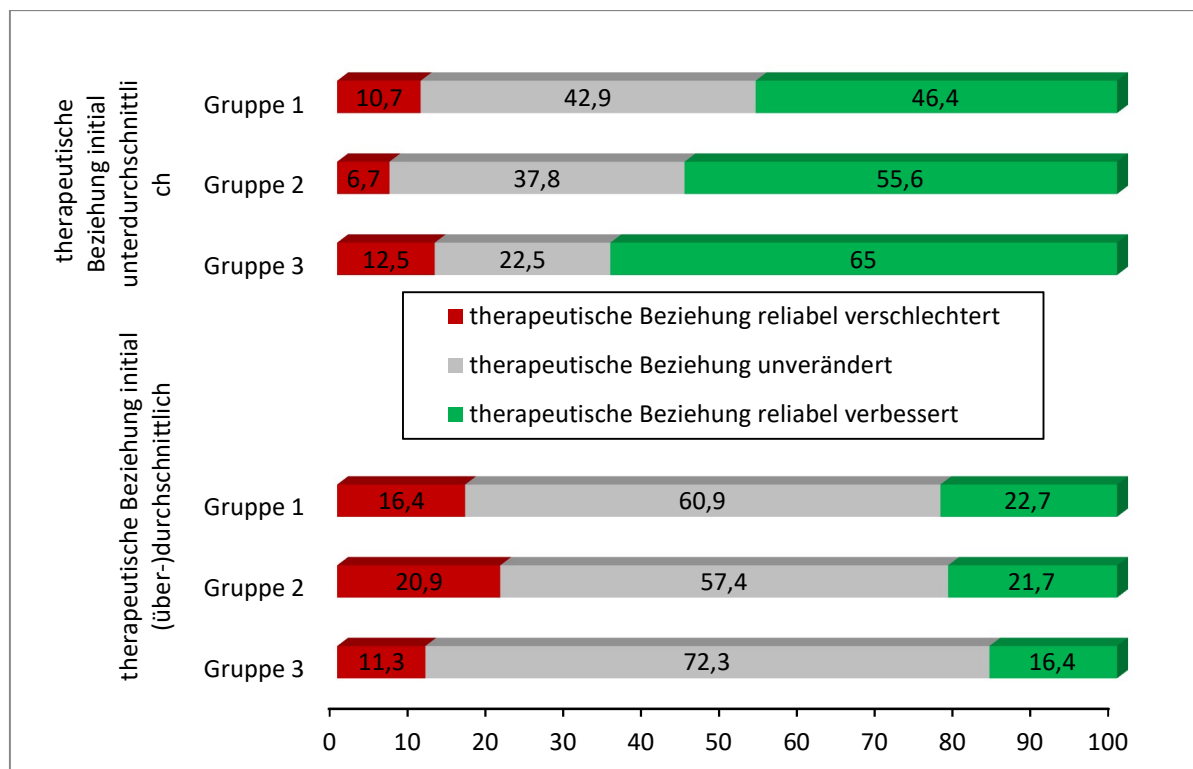


Abbildung 14: Reliable Veränderungen der HAQ-Beziehungswerte zwischen Aufnahme und Entlassung (Prozentangaben) für Patienten mit initial unterdurchschnittlichen und (über-)durchschnittlichen Beziehungswerten.

Unterschiede bezüglich der abhängigen Variable „Reliable Veränderung der HAQ Beziehungsskala“ wurden durch ein gemischtes logistisches Regressionsmodell geprüft. Dabei wurden „keine Veränderung“ und „reliable Verschlechterung“ zusammengefasst und auf „0“ gesetzt, reliable Verbesserung wurde als „1“ kodiert. Entgegen der Hypothese zeigte sich kein signifikanter Interaktionseffekt zwischen „Gruppe“ und „HAQ-Beziehungsskala bei Aufnahme und/oder erster Zwischenerhebung <16. Perzentil“. Außerdem zeigten sich weder ein signifikanter Haupteffekt für „Gruppe“ noch signifikante Effekte für die Kovariaten „Aufenthaltsdauer“ und „GSI bei Aufnahme“. Der einzige signifikante Effekt zeigte sich für die Variable „HAQ-Beziehungsskala bei Aufnahme und/oder erster Zwischenerhebung <16. Perzentil“.

Tabelle 17: Gemischtes logistisches Regressionsmodell mit abhängiger Variable „reliable Veränderung Beziehungsskala HAQ“

	F-Wert (df1; df2)	p-Wert
Gruppe	0.23 (2; 21.16)	0.795
Th. Beziehung < 16. Perzentil	47.68 (1; 507)	< 0.0001
Gruppe* Th. Beziehung < 16.Perzentil	2.23 (2; 507)	0.108
Aufenthaltsdauer	1.98 (1; 507)	0.160
GSI Aufnahme	0.32 (1; 507)	0.573

## 4 DISKUSSION

Ziel der vorliegenden Arbeit war es, die Akzeptanz und Effektivität von Therapeutenfeedback zu unterschiedlichen Zeitpunkten im Therapieverlauf sowie von unterschiedlichen Feedbackinhalten – prozess- und/oder symptombezogen - zu untersuchen. Zu diesem Zweck wurden die Daten von 537 Patienten analysiert, deren Therapeuten einer von drei Untersuchungsgruppen (1.) symptombezogenes Feedback sechs Wochen nach Aufnahme, 2.) symptombezogenes Feedback nach Aufnahme, zwei und sechs Wochen nach Aufnahme, 3.) prozess- und symptombezogenes Feedback nach Aufnahme, zwei und sechs Wochen nach Aufnahme) zugeteilt wurden. Erwartet wurde ein besseres Therapieergebnis für Patienten mit einem frühen auffälligen Verlauf, deren Therapeuten von Therapiebeginn an Verlaufsfeedback erhalten hatten (Gruppe 2 und 3). Außerdem wurden bei Patienten der Gruppen 2 und 3 im Vergleich zu Patienten der Gruppe 1 längere Behandlungsdauern bei Patienten mit einem frühen auffälligen Verlauf und kürzere Behandlungsdauern bei Patienten, deren Symptomatik bereits zwei Wochen nach Aufnahme innerhalb des Normbereichs lag, erwartet. Schließlich wurde angenommen, dass Patienten mit initial unterdurchschnittlichen Werten auf der HAQ-Beziehungsskala, deren Therapeuten Rückmeldung über die therapeutische Beziehung erhalten hatten (Gruppe 3), bessere Beziehungswerte bei Entlassung und häufiger reliable Verbesserungen der Beziehungswerte aufwiesen als Patienten mit unterdurchschnittlichen initialen Beziehungswerten in den Gruppen 2 und 3. Die Ergebnisse der Studie werden im Folgenden diskutiert. Dabei wird zunächst auf die Repräsentativität der untersuchten Stichprobe sowie auf Unterschiede zwischen den Untersuchungsgruppen eingegangen. Anschließend werden die Ergebnisse der Hypothesenprüfungen zusammengefasst und bezugnehmend auf die Forschungsliteratur diskutiert. Außerdem werden Stärken und Limitationen des Studiendesigns betrachtet. Abschließend werden Schlussfolgerungen aus den Befunden der vorliegenden Arbeit gezogen und ein Ausblick auf mögliche zukünftige Anwendungen und Forschungsfragen gegeben.

### 4.1 Repräsentativität der Stichprobe

Die soziodemografischen und klinischen Merkmale der in dieser Arbeit untersuchten Stichprobe stimmen weitgehend mit Daten überein, die für stationäre Psychotherapiepatienten in Deutschland berichtet werden. Bezüglich Geschlechterverteilung, Durchschnittsalter, Bildungsstand und Beschäftigungsstatus ist die Stichprobe vergleichbar mit Angaben aus der Meta-Analyse von Liebherz und Rabung (2014), die Daten aus über 80 Stichproben aggregierten. Auch die Beeinträchtigungsschwere bei Behandlungsbeginn – sowohl bzgl. der psychischen Symptomatik (SCL) als auch der interpersonalen Probleme (IIP) – ist vergleichbar. Die durchschnittliche Behandlungsdauer von 97 Tagen liegt innerhalb des



Ranges (M +/- sd) der bei Liebherz und Rabung (2014) berichteten durchschnittlichen – und sehr heterogenen – Aufenthaltsdauern. Damit korrespondiert, dass die Mehrheit der in der Meta-Analyse untersuchten Stichproben aus Kliniken stammt, die einen psychodynamisch orientierten Therapieansatz verfolgen – ebenso wie die Einrichtung, in der die Stichprobe für diese Arbeit gewonnen wurde. Im Unterschied zu Liebherz und Rabung (2014) waren die Patienten der hier untersuchten Stichprobe häufiger alleinstehend (und seltener verheiratet) und bekamen als Erstdiagnose etwas häufiger eine F3-Diagnose (Affektive Erkrankungen) und etwas seltener eine F6-Diagnose. Außerdem war in der hier untersuchten Stichprobe die Komorbidität mit 93,4% höher.

Die durchschnittliche Effektstärke (Aufnahme – Entlassung) im GSI liegt in der vorliegenden Studie mit  $d=0,72$  höher als die in der Meta-Analyse von Steffanowski, Löschmann, Schmidt, Wittmann und Nübling (2007) für die psychosomatische Rehabilitation ermittelte durchschnittliche Effektstärke von  $d=0,59$ . Allerdings liegt sie geringfügig unter Effektstärken, die für den GSI in psychosomatischen Krankenhausbehandlungen ( $d=0,79$  bei Probst et al., 2009,  $d=0,83$  bei Tritt et al., 2003), in psychoanalytisch orientierter stationärer Therapie (Franz et al., 2000:  $d=0,84$ ) und in derselben Einrichtung wie in vorliegender Studie (Dinger, Strack, Leichsenring, Wilmers & Schauenburg, 2008,  $d=0,80$ ) berichtet werden. Dabei könnten Selektionseffekte eine Rolle spielen – z.B. wurden in der hier untersuchten Stichprobe Patienten mit geplanter Kurzzeittherapie ausgeschlossen, was sich z.B. im Vergleich zu Probst et al. (2009) in höheren initialen Beeinträchtigungswerten (GSI) und einem höheren Ausmaß an Ko-/Multimorbidität wieder spiegelte.

## 4.2 Vergleichbarkeit der Untersuchungsgruppen

Patienten der Gruppe 3 waren älter als Patienten der anderen beiden Gruppen, lebten häufiger in fester Partnerschaft und hatten ein höheres formales Bildungsniveau. Damit stellten sich die drei Untersuchungsgruppen zwar in vielen, jedoch nicht in allen erfassten soziodemografischen Variablen als vergleichbar dar. Da soziodemografische Faktoren jedoch einen vernachlässigbar geringen Einfluss auf den Behandlungsverlauf und den Therapieerfolg haben (zusammenfassend Bohart & GreavesWade, 2013), wurde in den Analysen nicht für diese drei Variablen kontrolliert.

Bezüglich klinischer Variablen zeigten sich Gruppenunterschiede bei den Diagnosen: In Gruppe 3 wurden mehr F4-Diagnosen und weniger F3-Diagnosen vergeben als in den anderen beiden Gruppen, außerdem war die durchschnittliche Anzahl an Diagnosen zum Aufnahmezeitpunkt und somit die Ko-/Multimorbidität geringer. Da sich diese Unterschiede jedoch nicht auf die psychometrischen Beeinträchtigungsmaße auswirkten – die Gruppen unterschieden sich bei Aufnahme weder im Ausmaß der psychischen Beeinträchtigung (patientenseitig: GSI, therapeutenseitig: Beeinträchtigungsschwerescore BSS), noch in der

interpersonalen Beeinträchtigung (IIP Gesamtwert) -, wurde der Unterschied in den Diagnosekategorien F3 und F4 nicht als bedeutsam für Behandlungsverlauf und -erfolg eingeschätzt.

In der Einschätzung der therapeutischen Arbeitsbeziehung gab es keine a priori Unterschiede zwischen den Gruppen, was insbesondere für Fragestellung 4 bedeutsam war.

### 4.3 Zusammenfassung und Diskussion der Ergebnisse (Hypothesenprüfung)

Im Folgenden werden die Ergebnisse der vier Fragestellungen (vgl. 1.9) zusammengefasst und diskutiert.

#### 4.3.1 Fragestellung 1: Praktikabilität, Akzeptanz und Nutzen des computergestützten Therapeutenfeedbacks

##### **4.3.1.1 Praktikabilität**

In der vorliegenden Studie lagen von 619 Patienten mit kompletten Datensätzen zu Aufnahme und Entlassung für 523 Patienten auch komplette Verlaufsdaten vor, was einem Anteil von 15,5% fehlender Werte bei den Verlaufserhebungen entspricht. Damit wurden in etwa so gute „Ausfüllraten“ erzielt wie bei Puschner et al. (2009), die eine Rate von 80% kompletten Befragungen berichten – ebenfalls im stationären Setting mit deutlich beeinträchtigten Patienten, allerdings unterstützt durch ausschließlich für die Datenerhebung zuständiges Studienpersonal. Das Ergebnis dieser Studie zeigt, dass Verlaufserhebungen bei Patienten unter Routinebedingungen auch ohne beträchtlichen finanziellen und personellen Aufwand durchführbar sind. Dies war hier sogar bei Anwendung umfangreicher psychometrischer Instrumente und der Befragung eines heterogenen, unselektierten psychiatrischen Patientenlientels der Fall. Insbesondere Feedbackstudien, die in der ambulanten Psychotherapie durchgeführt wurden, berichten hohe Drop-out Raten sowie einen beträchtlichen Anteil an fehlenden Datensätzen (z.B. deJong et al., 2014; Scheidt et al., 2012). Das stationäre Setting mit seinen Routinen und regulierten Abläufen kann die sehr gute „Ausfüllquote“ begünstigt haben. Der geringe Anteil an fehlenden Daten kann auch dadurch beeinflusst worden sein, dass die Einrichtung im Rahmen ihres QS-Systems bereits jahrelange Erfahrung mit der systematischen Erhebung von Patientendaten im Routinebetrieb hatte.

Nicht nur die Patientenerhebungen im Therapieverlauf erwiesen sich als praktikabel, auch das computergestützte Feedback an die Therapeuten war im Routinebetrieb gut zu implementieren. Das Kriterium für die Praktikabilität des Monitoring- und Feedbacksystems Web-AKQUASI und die protokollkonforme Umsetzung der Intervention wurde mit dem Vorliegen eines beantworteten Therapeuten-Feedbackbogens streng gewählt. Da für 94,1% aller Feedbacknachrichten nach Aufnahme und im Therapieverlauf ausgefüllte Therapeuten-Feedbackbögen vorlagen, kann davon ausgegangen werden, dass Web-AKQUASI für die

Behandler sehr gut zu bedienen war und eine hohe Adhärenz bezüglich der Feedbackintervention erreicht werden konnte. Dies ist erstaunlich, da keine Maßnahmen ergriffen worden waren, um die Adhärenz zu stärken. So wurde weder das Ausfüllen der Therapeuten-Feedbackbögen kontrolliert, noch wurden computergestützte Reminder geschaltet oder spezielle Veranstaltungen wie z.B. Qualitätszirkel durchgeführt, um die Intervention zu stärken. Weder Patienten noch Therapeuten wurden bei nicht beantworteten Fragebögen mit Nachfragen oder Erinnerungen „konfrontiert“.

Bei der Entlassungserhebung wurden nur noch für 68,5% der Feedbacknachrichten von den Behandlern Rückmeldebögen ausgefüllt. Dies muss jedoch nicht bedeuten, dass knapp ein Drittel der Rückmeldungen zum Behandlungsende – die das Feedback über den gesamten Therapieverlauf und das Behandlungsergebnis enthalten – von den Therapeuten nicht betrachtet wurden. Bei der Vorstellung der Studie war die Überprüfung der Effektivität der verschiedenen Verlaufsrückmeldungen als Studienziel betont worden sowie die Bedeutung, die den Therapeuten-Feedbackbögen im Behandlungsverlauf zukommt. Die Rückmeldung zum Entlassungszeitpunkt war nicht mehr Bestandteil der Intervention, was bei der Vorstellung der Studie entsprechend kommuniziert worden war. Daher wird angenommen, dass einige Behandler zwar die Ergebnissrückmeldung bei Entlassung angesehen hatten, das Ausfüllen des Feedbackbogens jedoch nicht mehr als notwendig angesehen haben.

Voraussetzung für praktikable Verlaufserhebungen und -rückmeldungen im Routinebetrieb scheint ein computergestütztes Datenerhebungs- und Feedbacksystem zu sein. So berichten Berking et al. (2006), dass das Feedback trotz des in ihrer Studie gefundenen positiven Effekts aufgrund des hohen Aufwands der Papier-Bleistift-Erhebung nicht in die Klinikroutine implementiert werden konnte. Die zuletzt aus Deutschland publizierte Feedbackstudie - ebenfalls in der stationären Versorgung durchgeführt (Probst et al., 2013, 2014) - verwendete zwar das Programm OQ-Analyst für die Datenauswertung, stützte sich jedoch sowohl bei der Datenerhebung als auch bei der Übermittlung des Feedbacks auf die Papier-Bleistift-Methode. Auch hier wurde das Verlaufsfeedback ausschließlich im Rahmen der Studie durchgeführt und danach eingestellt. Daher sollte bereits bei der Planung von praxisnahen patientenzentrierten Studien die spätere Umsetzbarkeit der Intervention in den Routinebetrieb mit bedacht werden.

#### **4.3.1.2 Akzeptanz und Nutzen**

In der vorliegenden Studie stimmten die Behandler der überwiegenden Mehrheit der Feedbacknachrichten zu. Während zwischen den Gruppen 2 und 3 keine Unterschiede in den „Zustimmungsraten“ festzustellen waren, war die Zustimmung in Gruppe 1 signifikant geringer. Dies spricht dafür, dass Feedback etwas besser akzeptiert wird, wenn es kontinuierlich bzw. von Therapiebeginn an gegeben wird und nicht erst zu späteren Zeitpunkten im Behandlungsverlauf. Im Vergleich zu Gruppe 2 wurden in Gruppe 3, die auch Rückmeldung

über die therapeutische Beziehung erhalten hatte, die Feedbacknachrichten seltener als (reine) Bestätigung und häufiger als Ergänzung des klinischen Eindrucks bewertet. Bezüglich des Nutzens für die Behandlung äußerten sich die Therapeuten zurückhaltend: Die Informationen wurden in allen drei Untersuchungsgruppen am häufigsten als „etwas relevant“ und das Feedback meistens als „etwas hilfreich“ für die weitere Behandlung bewertet. Dabei bewerteten Therapeuten in Gruppe 3 das Feedback als hilfreicher und den Inhalt als relevanter im Vergleich zu Therapeuten der Gruppen 1 und 2, die ausschließlich symptombezogenes Feedback erhalten hatten.

Obwohl mittlerweile eine Reihe von Studien zur Effektivität von Verlaufsfeedback vorliegen, wurde bisher nur selten untersucht, wie Therapeuten das Feedback bewerten, ob und in welcher Form sie es für hilfreich halten und wie sie es für ihre Arbeit nutzen. Das Feedback wurde von den hier untersuchten Therapeuten weniger positiv wahrgenommen als in einer norwegischen Feedback-Studie (Amble, Gude, Ulvenes, Stubdal, & Wampold, 2016), in der der „Gesamtnutzen“ des Feedbacks (basierend auf OQ-Analyst) mit 4.00 bewertet worden war (1: äußerst gering bis 5: sehr stark). Da sich das Feedback von OQ-Analyst und Web-AKQUASI – obwohl auf unterschiedlichen Modellen basierend (vgl. 1.4) - in der Darstellung sehr ähnelt, werden die Gestaltung/Präsentation der Rückmeldungen oder ihr Inhalt als mögliche Gründe für unwahrscheinlich gehalten. Auch in der bisher einzigen Studie aus der Lambert-Gruppe, die Akzeptanz und Nutzen des Feedbacks aus Therapeutensicht erfragt hatte (Lambert, Whipple et al., 2001), wurde das Verlaufsfeedback positiver bewertet als in der vorliegenden Studie (Die Hälfte von 30 befragten Therapeuten bewerteten das Feedback als „fast immer“ oder „häufig“ hilfreich). Da retrospektiv eine globale Einschätzung abgegeben wurde und nicht wie in dieser Arbeit jede Feedbacknachricht bezüglich ihres Nutzens bewertet worden war, sind diese Angaben allerdings vorsichtig zu interpretieren. Möglich ist eine geringerer „Enthusiasmus“ in der hier untersuchten Therapeutenstichprobe aufgrund einer längeren Gewöhnung an Feedback: Im Unterschied zu Amble et al. (2016) und Lambert, Whipple et al. (2001), wo vor den berichteten Studien keine Feedbackinterventionen durchgeführt worden waren, war hier das Verlaufsfeedback sechs Wochen nach Therapiebeginn bereits seit einiger Zeit implementiert, Ergebnisfeedback war bereits jahrelange Routine.

Mittlerweile werden in Studien, die sich auf das Design der Lambert-Arbeitsgruppe stützen, häufig „clinical support tools“ (CST) verwendet, die die Behandler mit Hinweisen zur Problemlösung in der therapeutischen Arbeit mit NOT-Patienten unterstützen sollen. Allerdings wurde bisher nicht untersucht bzw. berichtet, wie Therapeuten deren Nutzen bewerten. Bei ausschließlichem Prozessfeedback (McClintock et al., 2017) zeigten sich die Therapeuten sehr zufrieden mit dem Feedbacksystem (Mittelwert 4.17 bei Likert-Skala – 5: völlig zufrieden) und bewerteten das Feedback außerdem als hilfreich bis sehr hilfreich

(Mittelwert 4.0 bei Likert-Skala – 5: sehr hilfreich). In der vorliegenden Studie bewerteten die Therapeuten, die auch Prozessfeedback erhalten hatten (Gruppe 3) das Feedback als hilfreicher und den Inhalt als relevanter im Vergleich zu Therapeuten der Gruppen 1 und 2 mit ausschließlich symptombezogenem Feedback. Dieses Ergebnis unterstützt den Befund, dass Feedback mit größerer Wahrscheinlichkeit akzeptiert wird, wenn es vom Empfänger als persönlich relevant angesehen wird („Perceived validity“, Claiborn & Goodyear, 2005) sowie die Annahme, dass Behandler Prozessfeedback gegenüber aufgeschlossener sein könnten, da es einfacher in therapeutisches Handeln umgesetzt werden kann (Mc Clintock et al. 2017). Zudem wird diskutiert, dass Therapeuten (ausschließliches) Outcome-Feedback als bewertend und potenziell bedrohlich wahrnehmen können (Boswell, Krauss, Miller, & Lambert, 2015). Allerdings waren die Bewertungen im Vergleich zu McClintock et al. (2017) weniger positiv: Die überwiegende Mehrheit der Rückmeldungen wurde auch in Gruppe 3 als lediglich „etwas hilfreich“ und der Inhalt der Feedbacknachrichten als „etwas relevant“ beurteilt. Obwohl in Gruppe 3 die Skala zur therapeutischen Beziehung an den Anfang der Feedbacknachricht gestellt worden war, könnte das Feedback hier von den Symptomskalen – allein aufgrund der zahlenmäßigen „Überlegenheit“ (eine Prozess-Skala gegenüber 10 Symptomskalen) – dominiert worden sein.

#### 4.3.2 Fragestellung 2: Vergleich des Therapie-Outcomes

Wie erwartet, unterschieden sich die Gruppen nicht bezüglich der Outcome-Variablen GSI bei Entlassung, reliable Veränderung des GSI und des Auffälligkeitssignals. Der in den Hypothesen postulierte Interaktionseffekt zwischen Gruppe und Auffälligkeitssignal im frühen Behandlungsverlauf konnte für zwei der drei verwendeten Outcome-Maße - GSI bei Entlassung und das Auffälligkeitssignal - bestätigt werden. Dies bedeutet, dass Patienten, die zwei Wochen nach Behandlungsbeginn als auffällig klassifiziert wurden, signifikant häufiger ein gutes Behandlungsergebnis (niedrigere GSI-Werte, geringerer Anteil an auffälligen Verläufen) erzielten, wenn die Therapeuten im frühen Therapieverlauf Feedback erhalten hatten. Die Effektivität von Verlaufsfeedback in der Psychotherapie wird aktuell als deutlich zurückhaltender bewertet als in der ersten Meta-Analyse von Shimowaka et al. (2010), bis hin zu Befunden, die die Effekte als minimal bis nicht vorhanden bewerten (Lutz et al., 2012; Kendrick et al., 2016). In der vorliegenden Studie erscheint es noch erstaunlicher, dass ein Effekt (für zwei Outcome-Maße) gefunden wurde, da nicht nur „aktive“ Behandlungsformen verglichen wurden, sondern sogar ausschließlich Bedingungen, die *alle* ein Verlaufsfeedback beinhalteten – was in den bisher publizierten Feedbackstudien nie der Fall war.

In den Einzelvergleichen zeigten sich signifikante Unterschiede im GSI bei Entlassung und im Auffälligkeitssignal zwischen „früh auffälligen“ Fällen der Gruppen 2 und 1, nicht jedoch zwischen den Gruppen 3 und 1. Das Feedback über die therapeutische Beziehung ab

Behandlungsbeginn hatte also keinen zusätzlichen positiven Effekt auf das Behandlungsergebnis. Gleichzeitig kann nicht gefolgert werden, dass das Feedback in Gruppe 3 „weniger erfolgreich“ war als in Gruppe 2, da sich diese beiden Feedbackgruppen nicht bezüglich der Outcome-Maße – auch nicht für die Subgruppe der „früh auffälligen“ Verläufe – unterschieden.

Der postulierte Interaktionseffekt zeigte sich nicht für die Outcome-Variable „Reliable Veränderung“, auch wenn die Betrachtung der Prozentsätze dies hätte vermuten lassen (43% reliable Verbesserungen bei früh auffälligen Patienten in Gruppe 1 vs. 61% reliable Verbesserungen in Gruppe 2 und 64% in Gruppe 3). Dies zeigt, dass die Wahl des Outcome-Kriteriums und der verwendeten Methode zur Ergebnismessung bedeutsam ist. Hier handelt es sich mit dem GSI sogar um dieselbe Datenquelle - es wurde durch Verwendung des „Reliable Change Index“ lediglich ein indirektes Veränderungsmaß gewählt. Auch weitere Studien berichten signifikante Feedbackeffekte für den Wert des Outcome-Instruments bei Entlassung, jedoch nicht für den RCI (Delgadillo et al., 2018). Dies kann damit zusammenhängen, dass beim RCI eine Konfundierung des Outcome-Maßes mit einer Kovariate besteht, da die Beeinträchtigungsschwere zu Behandlungsbeginn im Outcome-Maß RCI enthalten ist. Da in der Psychotherapieforschung nicht durchgehend einheitliche Ergebnismaße und Methoden zur Ergebnismessung verwendet werden (Stieglitz & Hiller 2014), empfiehlt sich für eine Vergleichbarkeit der Studien die Darstellung verschiedener, häufig verwendeter Outcome-Maße und –methoden.

Da sich in der vorliegenden Arbeit für Patienten mit frühem auffälligen Verlauf bezüglich zweier Outcome-Maße ein positiver Effekt von zusätzlichem Feedback nach Aufnahme und zwei Wochen danach zeigte (Gruppe 2), wird gefolgert, dass bei der Implementierung von Verlaufsfeedback in die psychotherapeutische Versorgung auf Feedback im frühen Therapieverlauf nicht verzichtet werden sollte. Dies ist allerdings bei Verlaufsfeedback nach den unter 1.4.2 dargestellten Bewertungsalgorithmen (z.B. OQ-Analyst) nicht möglich, da mindestens drei Patientenerhebungen benötigt werden, um den individuellen Symptomverlauf mit erwarteten Verlaufskurven abzugleichen. Dadurch ist bei Verwendung dieses Algorithmus in Fällen mit festgelegten kurzen Behandlungsdauern (z.B. Newnham et al., 2010) Feedback trotz sehr engmaschiger Erhebungen erst nach der Hälfte der Therapie möglich oder es werden sogar Fälle mit weniger als drei Erhebungen von den Analysen ausgeschlossen (z.B. deJong et al., 2014; Probst et al., 2013, 2014).

Es wird berichtet, dass Verlaufsfeedback nicht geeignet sei, auffällige Verläufe zu reduzieren, sondern nur weitere Verschlechterungen aufzuhalten oder abzumildern (Amble et al., 2014; Probst et al., 2013; de Jong et al., 2014). Dies war in der vorliegenden Studie nicht der Fall: Hätten sich Patienten mit frühem auffälligem Behandlungsverlauf in den Gruppen 2 und 3 nur weniger verschlechtert als „früh auffällige“ Patienten in Gruppe 1, dann hätte es zwar durchaus

positive Effekte bzgl. des Outcome-Maßes „GSI-Wert bei Entlassung“ geben können, jedoch nicht bzgl. des Auffälligkeitssignals bei Entlassung. Letzteres bewertet das Behandlungsergebnis nur dann als positiv (und nicht auffällig), wenn signifikante Verbesserungen in der Mehrzahl der erhobenen Skalen gemessen wurden (vgl. 2.6.1). Künftige Studien sollten daher nicht ausschließlich auf Effektgrößen der Ergebnismaße fokussieren, sondern untersuchen, ob Feedback helfen kann, Patienten mit ungünstigem Behandlungsverlauf im Vergleich zu auffälligen Patienten ohne Therapeutenfeedback in einem signifikant gebessertem Gesundheitszustand zu entlassen (und nicht nur in einem weniger verschlechterten Zustand).

Da das Feedback in Gruppe 3, wie unter 4.3.1.2 dargestellt, von den Behandlern etwas positiver bewertet worden war als in den anderen beiden Gruppen, hätte ein (kleiner) zusätzlicher, positiver Effekt auf das Therapie-Outcome angenommen werden können, da Feedback dann zu Verhaltensänderungen führt, wenn es wahrgenommen und als sinnvoll beurteilt wird (Riemer & Bickman, 2011). Dass sich das zusätzliche Prozessfeedback nicht positiv auf das Behandlungsergebnis ausgewirkt hat, könnte mit Deckeneffekten bei der Messung der therapeutischen Beziehung zusammenhängen. Auch die bereits erwähnte „Dominanz“ der Symptomskalen bei der Feedbacknachricht könnte eine Rolle spielen, so dass das Prozessmaß nicht so intensiv betrachtet und reflektiert wurde, wie es vielleicht bei ausschließlichem Prozessfeedback der Fall gewesen sein wäre.

#### 4.3.3 Fragestellung 3: Vergleich der Behandlungsdauer

Die Hypothesen bezüglich des Effekts von Feedback in frühem Behandlungsverlauf auf die Behandlungsdauer im Sinne einer Steigerung der Effizienz – längere Behandlungsdauer für Patienten mit frühem auffälligen Verlauf, kürzere Behandlungsdauern für Patienten, deren Symptomwerte bereits früh im „gesunden“ Normbereich liegen – konnten nicht bestätigt werden. Zwar zeigte sich eine differenzielle Allokation der Behandlungsdauer, da bei der ersten Zwischenerhebung auffällige Patienten signifikant länger behandelt wurden als nicht auffällige Patienten. Gleichzeitig wurden Patienten, deren Symptomatik bei der ersten Zwischenerhebung im Normbereich lag, signifikant kürzer behandelt als Patienten mit „krankheitswertiger“ Symptomatik bei der ersten Zwischenerhebung. Allerdings war dies bei allen Gruppen der Fall und zeigte sich nicht in besonderem Maße für die Gruppen mit frühem Verlaufsfeedback. Damit konnte ein Befund, mit dem die Wirkweise des Verlaufsfeedbacks in einigen Studien erklärt wurde – die stärkere differenzielle Allokation des Behandlungsumfangs an Beeinträchtigungsschwere und Gesundheitsverlauf (vgl. 1.5.4.2) – nicht repliziert werden. Als Erklärungsversuch erscheinen Merkmale des Settings und der Einrichtung plausibel: Mit knapp 14 Wochen war die durchschnittliche Behandlungsdauer bereits sehr lang – nicht untypisch für stationäre, überwiegend psychodynamisch orientierte Psychotherapie in

Deutschland (vgl. Liebherz & Rabung, 2014), jedoch überhaupt nicht vergleichbar mit der Mehrzahl an publizierten Feedback-Studien, die in einem Setting ambulanter „Ultra-Kurzzeittherapie“ durchgeführt wurden, das im hiesigen Versorgungssystem vom Umfang her in etwa der Probatorikphase ambulanter Psychotherapie entsprechen würde. In Settings mit knappen Ressourcen – z.B. durchschnittlich knapp vier therapeutischen Sitzungen (Lambert, Whipple et al., 2001) - machen nur wenige zusätzliche Sitzungen für stärker beeinträchtigte NOT-Patienten einen merklichen und messbaren Unterschied. Ein Vorteil des Settings, in dem Lambert und Kollegen ihre Studien durchführen, ist die große Flexibilität der Behandler hinsichtlich der Anzahl an Sitzungen, die sie ihren Patienten zur Verfügung stellen können (Harmon et al., 2007). Im Unterschied dazu wurde in der vorliegenden Untersuchung allen Patienten ein vergleichsweise großzügiges Therapiekontingent zur Verfügung gestellt. Allerdings wäre eine weitere Ausdehnung der Behandlungsdauer für auffällige Patienten ggf. weder therapeutisch indiziert noch mit den Forderungen der Kostenträger vereinbar gewesen. Während sich die über die Gruppen hinweg vergleichbaren Behandlungsdauern von Patienten mit frühen auffälligen Verläufen erklären lassen, erscheint unklar, warum die frühe Rückmeldung eines Symptomprofils im „gesunden“ Normbereich keine Auswirkungen auf die Behandlungsdauer i.S. einer früheren Entlassung hatte. Denkbar ist, dass das stationäre Setting mit festgelegten Therapieabläufen weniger Flexibilität bietet, die Behandlungsdauer entsprechend des Gesundungsverlaufs der Patienten zu variieren. Hier scheint das ambulante Setting auch in Deutschland flexibler zu sein, wo trotz a priori Kontingentierungen nur etwa knapp drei Viertel der genehmigten Sitzungen ausgeschöpft werden (Gallas, Puschner, Kühn, & Kordy, 2010).

#### 4.3.4 Fragestellung 4: Vergleich der therapeutischen Beziehung

Die Hypothese, dass Therapeutenfeedback bezüglich der therapeutischen Beziehung (Gruppe 3) bei Patienten mit initial unterdurchschnittlichen Beziehungswerten zu besseren Werten der HAQ-Beziehungsskala bei Entlassung und einem größeren Ausmaß an reliabler Verbesserung der Beziehungswerte führt, konnte nicht bestätigt werden.

Die von McClintock et al. (2017) geäußerte Annahme “Outcome feedback may primarily affect outcome, and process feedback may primarily affect process“, p. 256), kann durch das Ergebnis der vorliegenden Untersuchung nicht untermauert werden.

Eine vergleichende Betrachtung mit Studien, die die „Clinical Support Tools“ (CST) angewandt hatten – die auch ein Instrument zur Erfassung der therapeutischen Beziehung enthalten – ist aufgrund des unterschiedlichen Designs schwierig. Zum einen werden CST ausschließlich bei der Minderheit der NOT-Fällen als zusätzliche Intervention eingesetzt: Die therapeutische Beziehung wird also nur dann erfasst, wenn der Symptomverlauf als „auffällig“ bewertet



worden war. Der zusätzliche kleine, aber signifikante Effekt der CST-Intervention auf das Behandlungsergebnis bei den Studien von Lambert und Kollegen könnte zudem auf ein „Design Artefakt“ zurück zu führen sein: Nach der Sitzung, in der das NOT-Alarmsignal generiert worden war, mussten Patienten in der CST-Bedingung noch mindestens drei weitere Sitzungen absolvieren, um in die Analyse eingeschlossen zu werden – Patienten in der Feedbackbedingung ohne CST mussten jedoch nur mindestens eine weitere Sitzung absolvieren (Harmon et al., 2007), so dass es bereits aus organisatorischen Gründen zu einer längeren Behandlungsdauer in der CST-Feedbackbedingung kam.

Möglicherweise sind bei prozessorientiertem Feedback – speziell wenn es auf die Wahrnehmung des Therapeuten durch den Patienten und Aspekte der therapeutischen Beziehung fokussiert – andere Aspekte von Bedeutung als bei ausschließlich symptomorientiertem Feedback. Beispielsweise könnte ein kombiniertes Patienten- und Therapeutenfeedback, das regelmäßig in der Therapie gemeinsam betrachtet, bewertet und bzgl. seiner Implikationen für den weiteren Behandlungsverlauf diskutiert wird, bei der Rückmeldung der therapeutischen Beziehung hilfreicher sein als bei ausschließlich symptomorientiertem Feedback. Flückiger et al. (2012) zeigten, dass Patienten, die explizit ermutigt werden, sich aktiv in die Behandlungsplanung einzubringen, bessere Beziehungswerte berichten als Patienten, die nicht derart ermutigt werden.

McClintock et al. (2017) berichten, dass die Feedbackintervention den Patienten in der ersten Therapiesitzung folgendermaßen vermittelt worden war: „Each of these components (i.E., outcome expectations, empathy, and alliance – *Anmerkung der Verfasserin*) is strongly related to treatment success, and so by maximizing these components in our treatment, we might be able to maximize your improvement as well“ (p. 251). Dieses Beispiel zeigt, dass bereits die Art der Patienteninformation einen Effekt auf die therapeutische Beziehung haben kann, wenn sie wie in diesem Beispiel auf Kollaboration fokussiert – dem „gemeinsam an einem Strang Ziehen“, das als wichtiger Bestandteil der therapeutischen Beziehung gilt (Flückiger, 2018).

#### 4.4 Stärken und Limitationen des Studiendesigns

##### 4.4.1 Zusätzliches Verlaufsfeedback als „Minimalintervention“

Die Intervention wurde als Versorgungsstudie mit hoher ökologischer Validität konzipiert. Die Therapeuten erhielten eine Einführung in die Bedienung des Programms Web-AKQUASI und eine schriftliche Information zur Interpretation der Rückmeldungen. Die Kontrolle, dass die Rückmeldungen wahrgenommen wurden, erfolgte lediglich über das Ausfüllen des Therapeuten-Feedbackbogens zu jeder Rückmeldung. Es wurde nicht kontrolliert, wie, wann oder wie lange sich die Therapeuten innerhalb des 7-Tage-Intervalls, das zum Ausfüllen der Feedbackbögen zur Verfügung stand, mit der Rückmeldung beschäftigten. Ferner fand keine Instruktion statt, welche Konsequenzen aus den Rückmeldungen gezogen werden sollten.

Letztere wurden zwar im Therapeuten-Feedbackbogen erfragt, jedoch nicht kontrolliert. Ob die Rückmeldung beachtet wurde und in welcher Form die Informationen genutzt wurden, war den Therapeuten komplett freigestellt. Durch stärkere Einbindung der Rückmeldungen in formale Abläufe (z.B. Diskussionen in Teambesprechungen, Qualitätszirkeln) hätte die Intervention möglicherweise gestärkt werden können. Die Intervention wurde allerdings bewusst so konzipiert, dass sie möglichst wenig zusätzliche Ressourcen in Anspruch nimmt und so gut wie möglich in den klinischen Alltag integrierbar ist, um Akzeptanz, Nutzen und Effektivität der Rückmeldungen unter Routinebedingungen zu untersuchen. Zudem scheinen Ausweitungen der Feedback-Intervention, z.B. durch zusätzliches Patientenfeedback oder speziell eingerichtete Qualitätszirkel, die Effektivität von Verlaufsfeedback nicht unbedingt zu erhöhen bzw. keine Garantie dafür zu sein, dass überhaupt ein Effekt festzustellen ist (z.B. Puschner et al., 2009).

#### 4.4.2 Zeitpunkte und Frequenz der Rückmeldungen

In den Studien, die positive Effekte therapiebegleitender Rückmeldungen zeigten (z.B. zusammenfassend Shimokawa et al., 2010) wurde vor jeder Therapiestunde (i.d.R. wöchentlich) der Gesundheitszustand des Patienten erfragt und zurückgemeldet. Es gibt Hinweise darauf, dass der Effekt des Feedbacks größer ist, je häufiger Therapeuten Feedbacknachrichten erhalten (Bickman et al., 2011). Allerdings kann Feedback auch bei nur einer Verlaufsmessung einen Effekt zeigen (Newnham et al., 2010).

Das Therapeutenfeedback war in dieser Untersuchung mit einer Rückmeldung nach der Aufnahme und zwei Rückmeldungen im Therapieverlauf nicht so engmaschig wie ursprünglich gewünscht, musste jedoch an die organisatorischen Gegebenheiten angepasst werden. Außerdem ist es als Schwäche anzusehen, dass es keine Gruppe ohne Verlaufsfeedback gab. Bei der Planung der Studie wurde ein Vergleich mit archivarischen Daten diskutiert: Es sollte eine Vergleichsstichprobe aus dem QS-Datensatz generiert werden aus der Zeit, bevor die Zwischenerhebung sechs Wochen nach Aufnahme eingeführt worden war. Allerdings war dieses Vorgehen aus datenschutzrechtlichen Gründen unzulässig: Die Patienten hatten damals ihr Einverständnis gegeben, dass ihre Daten zu Zwecken des internen Qualitätsmanagements ausgewertet werden, die Verwendung durch eine externe Stelle zu Forschungszwecken war nicht vorgesehen.

#### 4.4.3 Cluster-Randomisierung

In dieser Studie wurden die Behandler zufällig einer der drei Untersuchungsgruppen zugeteilt. Es erfolgte nicht, wie bei der Mehrzahl der Feedback-Studien (z.B. allen Studien der Lambert-Arbeitsgruppe, Lambert, 2015), eine Randomisierung nach Patient. Es wurde entschieden, Therapeuten statt Patienten zu randomisieren, um eine eventuelle „Kontaminierung“ des

Effekts zu vermeiden. Dies geschah in Übereinstimmung mit dem in Therapiestudien üblichen Vorgehen, dass Behandler in Therapiestudien üblicherweise ausschließlich innerhalb einer treatment-Bedingung agieren (Gilbody, House & Sheldon, 2004). Falls auf Ebene der Patienten randomisiert worden wäre, hätte ein Therapeut z.B. für manche Patienten Rückmeldung über die therapeutische Beziehung erhalten und für andere nicht. Dies hätte zu einer Sensibilisierung bezüglich der therapeutischen Allianz bei allen behandelten Patienten führen können, so dass der „wahre“ Effekt der Intervention nicht mehr hätte entdeckt werden können.

Angesichts von Befunden zur unterschiedlichen Effektivität von Therapeuten (Okiishi, Lambert, Nielsen & Ogles, 2003; Wampold & Brown, 2005) ist es allerdings nicht auszuschließen, dass aufgrund der Cluster-Randomisierung Unterschiede zwischen Therapeuten den Effekt der Feedbackintervention kontaminiert haben könnten, zumal die Stichprobengröße mit 21 Therapeuten gering ist. Es wurde versucht, einen systematischen Selektionseffekt zu vermeiden, indem die Therapeuten zufällig eine der drei Gruppen zugewiesen wurden und versucht wurde, Settingvariablen konstant zu halten (Stratifizierung nach Behandlungsschwerpunkten).

Durch die Cluster-Randomisierung können außerdem Probleme bei einem Therapeutenwechsel entstehen, da dadurch ggf. die ursprüngliche Gruppenzuweisung nicht mehr einzuhalten ist. Ein Wechsel zu einem Therapeuten einer anderen Untersuchungsgruppe fand in der vorliegenden Untersuchung allerdings nur bei 29 Patienten der ITT-Stichprobe statt.

#### 4.4.4 Auswahl der Erhebungsinstrumente

Mit der SCL-90-R wurde eines der gängigsten störungsübergreifenden Instrumente zur Erfassung der psychischen Beeinträchtigung, mit dem IIP der am häufigsten zur Bestimmung interpersonaler Probleme eingesetzte Selbstbeurteilungsbogen verwendet. Allerdings sind beide Instrumente sehr umfangreich; ein häufiger Einsatz (z.B. im wöchentlichen Rhythmus) erscheint wenig praktikabel. Die Verwendung im Behandlungsverlauf ist überwiegend der Tatsache geschuldet, dass beide Instrumente bereits im Rahmen der QS eingesetzt worden waren (auch bei der bereits implementierten Zwischenerhebung sechs Wochen nach Aufnahme). Kurzformen von etablierten Ergebnismaßen sind ökonomischer und erlauben eine häufigere Messung im Therapieverlauf (Lutz et al., 2006). Solche Kurzformen teilen jedoch mit den hier verwendeten Langformen das Problem, dass sie zur Therapieevaluation und nicht zur Veränderungsmessung im Therapieprozess entwickelt wurden, so dass die Veränderungssensitivität häufig vernachlässigt und die Frage der Behandlungssensitivität – der Unterscheidungsfähigkeit zwischen wirksamer und unwirksamer Therapie – meist gar nicht adressiert wird (Stieglitz & Hiller, 2014).

Es gilt als generelles Problem, dass speziell für die Veränderungsdiagnostik entwickelte Verfahren fehlen und Tests in der Regel für die Erst- und Einmaluntersuchung entwickelt werden (Stieglitz & Hiller, 2014; Ogles, 2013). Der OQ-45 wurde zwar speziell für die Verlaufsmessung entwickelt (Lambert et al., 1996), allerdings verwenden die meisten Feedbackstudien, die den OQ-45 einsetzen, diesen gleichzeitig als alleiniges Outcome-Kriterium – das heißt, unabhängige Variable und abhängige Variable werden durch dasselbe Instrument erfasst (z.B. Harmon et al., 2007; Lambert, Whipple et al., 20; Whipple et al., 2003). Studien, die neben dem Instrument zur Verlaufsmessung noch andere Ergebnismaße einsetzen, sind selten (z.B. Newnham et al., 2010). In der vorliegenden Studie wurde mit dem Auffälligkeitssignal ein multidimensionales Ergebnismaß gewählt, in das neben den Veränderungswerten auf den zur Verlaufsmessung eingesetzten Instrumenten SCL-90-R und IIP noch weitere Maße – sowohl aus Patienten- als auch Therapeutesicht, mit direkter und indirekter Veränderungsmessung – eingingen. Damit ist das hier verwendete zentrale Ergebnismaß elaborierter als Outcome-Maße anderer Feedbackstudien, was mit der ursprünglichen Zielsetzung bei der Entwicklung des hier verwendeten Systems Web-AKQUASI –dem Einsatz in der Qualitätssicherung (vgl. 1.2) – zusammenhängt.

#### 4.4.5 Feedback ausschließlich an Einzeltherapeuten

Im hier untersuchten stationären Setting stellt die Einzelpsychotherapie nur eine Behandlungskomponente eines multimodalen Behandlungsprogramms dar, das von verschiedenen Einzel- und gruppentherapeutischen Angeboten, Gesprächen mit dem Pflorgeteam und dem Einfluss der Mitpatienten geprägt ist. Schauenburg, Dinger und Strack (2005) untersuchten auf einer Psychotherapiestation - wie in der vorliegenden Untersuchung psychodynamisch ausgerichtet - den Einfluss des Einzeltherapeuten auf das Behandlungsergebnis. Es konnte gezeigt werden, dass auch bei dem für die stationäre Psychotherapie typischen integrierten Behandlungskonzept der Einfluss des Einzeltherapeuten bestehen bleibt (10% aufgeklärte Varianz bezüglich Therapieerfolg). Dieser Befund wird gestützt durch Studien, die positive Effekte von Verlaufsfeedback im stationären Setting berichten (Berking et al., 2006; Newnham et al., 2010; Simon et al., 2013). Berking et al. (2006) argumentieren sogar, dass Verlaufsfeedback in der stationären Psychotherapie eine stärkere Wirkung haben könnte - z.B. aufgrund größerer therapeutischer Einflussmöglichkeiten als in ambulanter Behandlung. Puschner et al. (2009) dagegen – deren Untersuchung keinen Feedbackeffekt ergab -, nehmen an, dass in der stationären Versorgung ein stärkerer Fokus auf das Behandlungsteam als Empfänger der Rückmeldungen sinnvoll sein könnte. Auch Simon et al. (2013) vermuten einen größeren Effekt, wenn die Feedbackintervention dem Setting angepasst wird und z.B. Methoden integriert werden, um Verlaufsdaten mit allen an der Behandlung des Patienten beteiligten Personen zu

teilen. Die Untersuchung der Wirksamkeit von Verlaufsfeedback in der stationären Psychotherapie an alle Personen, die an der Behandlung des Patienten mitwirken, ist ein interessanter zukünftiger Forschungsgegenstand.

#### 4.5 Schlussfolgerungen

Trotz der diskutierten Schwächen des Designs der vorliegenden Studie – insbesondere eine fehlende Gruppe ohne Verlaufsfeedback und nur zwei zusätzliche Rückmeldungen in den Gruppen 2 und 3 - zeigte sich ein positiver Effekt der symptombezogenen Rückmeldungen im frühen Behandlungsverlauf bezüglich zweier Ergebnisparameter für Patienten mit einem frühen auffälligen Verlauf. Dies soll Einrichtungen oder Behandler ermutigen, Verlaufsmonitoring und –feedback auch dann zu implementieren, wenn eine sehr engmaschige (z.B. wöchentliche) Datenerhebung als inkompatibel mit der klinischen Routine oder als zu hoher organisatorischer Aufwand betrachtet wird.

Der Befund, dass Verlaufsfeedback nicht für alle Patienten wirksam ist, sondern nur für eine Subgruppe von Patienten mit auffälligem Verlauf („not on track“), konnte in dieser Untersuchung repliziert werden. Während sich die den Feedbacksystemen zugrunde liegenden Modelle zum Gesundungsverlauf und damit die Bewertungsalgorithmen unterscheiden (vgl. 1.4), erzielen sie überwiegend ähnliche Effekte: Bessere Ergebnisse für Patienten, die einen ungünstigen Behandlungsverlauf aufweisen. Es ist anzunehmen, dass das Auffälligkeitssignal – auch wenn es nach unterschiedlichen Algorithmen zustande kommt – ein wichtiger Wirkfaktor ist und zu einem erhöhten Aufmerksamkeitsfokus bei den Behandlern führt. Hier wäre es interessant, verschiedene Feedbackmodelle mit den jeweils unterschiedlichen Algorithmen zur Bestimmung eines „auffälligen“ Verlaufs gegeneinander zu testen – nicht nur hinsichtlich deren Effekte auf das Therapieergebnis, sondern auch bezüglich deren Auswirkungen auf den therapeutischen Prozess.

Damit Verlaufsfeedback eine Wirkung im Sinne einer Effizienzsteigerung entfalten kann, ist eine Flexibilität der Ressourcenallokation notwendig, so dass die Behandlungsdauer unabhängig von äußeren Restriktionen, z.B. Vorgaben der Kostenträger, an den individuellen Bedarf des einzelnen Patienten adaptiert werden kann. Versuche einer Effizienzsteigerung sollten nicht dazu führen, stark beeinträchtigte, sich nicht ausreichend oder nur sehr langsam verbessernde, zu Rezidiven neigenden Patienten die notwendigen Langzeittherapien zu versagen. Allerdings könnte es sinnvoll sein, Verlaufsfeedback in Systemen mit vergleichsweise großzügigen Ressourcen - wie dem hiesigen Versorgungssystem - stärker als bisher auf leicht beeinträchtigte Patienten oder Patienten mit schnellem Genesungsverlauf zu fokussieren. In Analogie zu den „Clinical support tools“, die Handlungsempfehlungen für stärker beeinträchtigte, nicht (ausreichend) respondierende Patienten geben, könnte daher ein

Tool nützlich sein, das die Behandler unterstützt, Therapieindikation und –ziele für Patienten mit „gesunden“ Normwerten zu überprüfen.

Die zusätzliche Rückmeldung der therapeutischen Beziehung aus Patientensicht hatte weder einen Effekt auf das Therapie-Outcome noch auf die Beziehungswerte bei Entlassung oder die Veränderung der therapeutischen Beziehung bei Patienten mit initial unterdurchschnittlichen Beziehungswerten. Für die Empfehlung, daher auf eine Verlaufsrückmeldung der therapeutischen Beziehung zu verzichten, erscheint die empirische Grundlage allerdings noch nicht fundiert genug. Beispielsweise sollte untersucht werden, ob bei Rückmeldungen zur therapeutischen Beziehung andere Wirkmechanismen zum Tragen kommen als bei symptomorientiertem Feedback und z.B. das Gespräch zwischen Patient und Therapeut über beziehungsorientierte Feedbackinhalte als Bestandteil der Intervention bedeutsam ist.

#### 4.6 Ausblick

In der Psychotherapieforschung wurde der patientenzentrierte Ansatz entwickelt, um über die Frage „What works“ hinaus zu differenzieren: „What works for whom under what conditions“. Diese Fragestellung sollte auch für die Untersuchung der Effektivität von Verlaufsfedback gelten. Nach vielen Replikationsstudien sollten zukünftig verstärkt differenzielle Effekte untersucht und die Frage adressiert werden, unter welchen Umständen (Setting, Versorgungsaspekte) und für welche Patienten welche Art von Feedback zu welchen Zeitpunkten und in welcher Frequenz geeignet ist, um Behandlungsergebnisse zu optimieren. Es wird vorgeschlagen, Rückmeldungen patientenspezifisch zu adaptieren und z.B. an unterschiedliche Verlaufsmuster anzupassen (Delgadillo, Moreea & Lutz, 2016).

Neben dem Monitoring des Behandlungsverlaufs eröffnen sich noch weitere Anwendungsgebiete, in denen eine kontinuierliche Erhebung und Rückmeldung von Patientendaten nützlich sein kann. Verlaufsmoitoring und -feedback wurden z.B. erfolgreich überprüft in der Rückfallprophylaxe nach stationärer Behandlung für Patienten mit Bulimia nervosa (Bauer, Okon, Meermann, & Kordy, 2012) und rezidivierender Depression (Kordy et al., 2016). Weiterhin bietet Verlaufsmoitoring und -feedback Potenzial für die Supervision (z.B. Swift et al. 2015). Ergebnismonitoring und Verlaufsfedback sind bisher nicht (oder selten) in die Ausbildungscurricula für Psychotherapeuten integriert. Die Integration von Verlaufsmoitoring und –feedback in die Psychotherapieausbildung würde auch dazu führen, zukünftig eine breitere Anwendung in der klinischen Praxis zu erzielen.

In den letzten Jahren wird zunehmend gefordert, dass die Psychotherapie eine Behandlungs- und Forschungskultur entwickeln soll, die Misserfolge und Nebenwirkungen systematisch erfasst (Strauß, Linden, Haupt, & Kaczmarek, 2012). Dabei mangelt es bisher an systematischen Konzeptionen, z.B. Definitionen für Misserfolg in der Psychotherapie, sowie

an schlüssigen Abgrenzungen von Misserfolg und therapeutischen Nebenwirkungen (Scott & Young, 2016; Linden, 2013). Verlaufsmonitoring und –feedback bieten hier die Möglichkeit, Auffälligkeiten im Behandlungsverlauf frühzeitig zu entdecken und eine „Fehlerkultur“ in die klinische Praxis zu implementieren, die nicht als (externe) Kontrolle, sondern als kontinuierlicher Verbesserungsprozess verstanden wird.

## 5 ZUSAMMENFASSUNG

Diese Arbeit untersuchte die Akzeptanz und Effektivität von symptom- und prozessbezogenem Therapeutenfeedback zu unterschiedlichen Zeitpunkten im Verlauf stationärer Psychotherapie. Psychometrische Daten aus Patientensicht wurden bei Aufnahme, zu zwei Zeitpunkten im Verlauf und bei Entlassung erhoben. Als Symptommaße wurden die Symptom-Checkliste (SCL-90-R) und das Inventar Interpersonaler Probleme (IIP) und als Prozessmaß der Helping Alliance Questionnaire (HAQ) eingesetzt. Datenerhebung, -verarbeitung und -rückmeldung erfolgten computer-gestützt mittels einer für die Studie adaptierten Version der Software Web-AKQUASI.

Es wurde ein cluster-randomisiertes Studiendesign gewählt, wobei 21 Bezugstherapeuten einer von drei Untersuchungsgruppen zugewiesen wurden:

- Gruppe 1: ausschließlich symptombezogenes Feedback (SCL-90-R, IIP), Rückmeldung 6 Wochen nach Aufnahme,
- Gruppe 2: ausschließlich symptombezogenes Feedback (SCL-90-R, IIP), Rückmeldung nach Aufnahme, 2 Wochen und 6 Wochen nach Aufnahme,
- Gruppe 3: symptom- und prozessbezogenes Feedback (HAQ, SCL-90-R, IIP), Rückmeldung nach Aufnahme, 2 Wochen und 6 Wochen nach Aufnahme,

Aufgrund der Vorbefunde wurden keine Haupteffekte für die Untersuchungsgruppen erwartet, sondern ein besseres Therapieergebnis für Patienten mit einem auffälligen Behandlungsverlauf bei der ersten Zwischenerhebung in den Gruppen 2 und 3 im Vergleich zu früh auffälligen Patienten der Gruppe 1. Außerdem wurde in den Gruppen 2 und 3 eine höhere Effizienz angenommen, die sich in einer längeren Behandlungsdauer von Patienten mit einem frühen auffälligen Verlauf und einer kürzeren Behandlungsdauer von Patienten, deren globaler Schwereindex GSI bei der ersten Zwischenerhebung im Normbereich lag, zeigen sollte. Außerdem wurde erwartet, dass Patienten mit initial unterdurchschnittlich bewerteter therapeutischer Beziehung in Gruppe 3 bessere Beziehungswerte bei Entlassung aufweisen als Patienten mit anfangs unterdurchschnittlichen Beziehungswerten der Gruppen 1 und 2.

Es wurden Daten von 537 Patienten (Intent-to-treat Stichprobe) analysiert. Aufgrund der hierarchischen Organisation der Daten (Untersuchungsebene: Patient, Randomisierungsebene: Therapeut) wurden die Hypothesen mit Hilfe von gemischten linearen Modellen (bei metrischen abhängigen Variablen) und gemischten logistischen Regressionsmodellen (bei binären abhängigen Variablen) geprüft.

Es zeigte sich eine sehr gute Umsetzbarkeit der Verlaufserhebungen und -rückmeldungen im klinischen Routinebetrieb. Die Adhärenz der Behandler, erfasst durch ausgefüllte Therapeuten-Feedbackbögen zu jeder Feedbacknachricht, war sehr gut, die inhaltliche



Relevanz des Feedbacks und der wahrgenommene Nutzen wurden jedoch als eher gering eingeschätzt. Therapeuten, die zusätzlich Feedback über die therapeutische Beziehung erhalten hatten, beurteilten die Rückmeldungen signifikant positiver als Therapeuten, die ausschließlich symptombezogenes Feedback erhalten hatten.

Bezüglich des Therapie-Outcomes zeigten sich wie erwartet keine Haupteffekte für die Untersuchungsgruppen. Es zeigten sich für zwei der drei gewählten Outcome-Maße (Globaler Schwere-Index GSI und Auffälligkeitssignal), nicht jedoch für die reliable Veränderung des GSI, signifikante Gruppenunterschiede bei Patienten mit einem frühen auffälligen Verlauf. In Einzelvergleichen ergab sich für Patienten mit einem frühen auffälligen Verlauf in Gruppe 2 eine signifikant niedrigere Symptombelastung (GSI) bei Entlassung und signifikant weniger auffällige (Gesamt-)Verläufe als bei früh auffälligen Patienten in Gruppe 1. Entgegen der Hypothese zeigte sich in den Einzelvergleichen jedoch kein signifikanter Unterschied in den Outcome-Maßen bei früh auffälligen Patienten der Gruppen 1 und 3. Damit konnte die Hypothese bezüglich des Therapie-Outcomes teilweise bestätigt werden. Die postulierte höhere Effizienz bei zusätzlichem „frühen“ Verlaufsfeedback (Gruppen 2 und 3) konnte nicht bestätigt werden: Patienten mit einem frühen auffälligen Verlauf wurden etwas länger und Patienten, deren GSI bei der ersten Zwischenerhebung im Normbereich lag, etwas kürzer behandelt. Dies war jedoch bei allen drei Untersuchungsgruppen gleichermaßen der Fall. Ein Effekt des prozessorientierten Feedbacks auf die Werte der HAQ-Beziehungsskala bei Entlassung oder auf die reliable Veränderung der Beziehungswerte konnte nicht festgestellt werden.

Aus den Ergebnissen wird gefolgert, dass symptombezogenes Feedback im frühen Behandlungsverlauf hilfreich ist, um für Patienten, die zu einem frühen Behandlungszeitpunkt als „auffällig“ bewertet werden, die Wahrscheinlichkeit für ein gutes Therapieergebnis zu erhöhen. Dieser Effekt zeigte sich bei im Vergleich zu TAU lediglich zwei zusätzlichen Rückmeldungen, so dass die Implementierung von Verlaufsmonitoring und -feedback auch empfohlen werden kann, wenn keine sehr engmaschigen Erhebungen möglich sind. Als Wirkfaktor wird das Alarmsignal, verbunden mit verstärkten Aufmerksamkeitsprozessen seitens der Behandler, angenommen. Die insgesamt sehr lange Behandlungsdauer bei gleichzeitig begrenzter Flexibilität werden als Gründe in Betracht gezogen, dass eine differenzielle Allokation der Behandlungsdauer als potenzieller weiterer Wirkfaktor des Verlaufsfeedbacks nicht bestätigt werden konnte. Eine Empfehlung, symptombezogenes Feedback um prozessorientierte Inhalte – hier die therapeutische Beziehung – zu ergänzen, kann auf Grundlage der Ergebnisse der vorliegenden Studie nicht gegeben werden. Allerdings sollte die empirische Basis bezüglich der Effekte prozessorientierter Verlaufsrückmeldungen noch erweitert werden.

## 6 LITERATURVERZEICHNIS

- Alexander, L. B., & Luborsky, L. (1986). The Penn Helping Alliance Scales. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook*. (pp. 325-366). New York: Guilford Press.
- Amble, I., Gude, T., Stubdal, S., Andersen, B. J., & Wampold, B. E. (2015). The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. *Psychother Res*, 25(6), 669-677. doi: 10.1080/10503307.2014.928756.
- Amble, I., Gude, T., Ulvenes, P., Stubdal, S., & Wampold, B. E. (2016). How and when feedback works in psychotherapy: Is it the signal? *Psychother Res*, 26(5), 545-555. doi: 10.1080/10503307.2015.1053552
- American Psychological Association: APA Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *Am Psychol*, 61(4), 271-285. doi: 10.1037/0003-066X.61.4.271
- Anderson, E. M., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *J Clin Psychol*, 57(7), 875-888. doi: 10.1002/jclp.1056
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve therapy outcomes: A randomized clinical trial in a naturalistic setting. *J Consult Clin Psychol*, 77(4), 693-704. doi: 10.1037/a0016062
- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: contrasting dose-effect and good-enough level models of change. *J Consult Clin Psychol*, 77(2), 203-211. doi: 10.1037/a0015235
- Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *J Consult Clin Psychol*, 68(6), 1027-1032. doi: 10.1037/0022-006x.68.6.1027
- Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *J Consult Clin Psychol*, 64(3), 619-622. doi: 10.1037/0022-006x.64.3.619
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcome benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *J Consult Clin Psychol*, 69(2), 184-196. doi: 10.1037/0022-006X.69.2.184
- Barkham, M., Stiles, W. B., & Shapiro, D. A. (1993). The shape of change in psychotherapy: Longitudinal assessment of personal problems. *J Consult Clin Psychol*, 61(4), 667-677. doi: 10.1037/0022-006X.61.4.667
- Barkham, M., Rees, A., Stiles, W. B., Shapiro, D. A., Hardy, G. E., & Reynolds, S. A. (1996). Dose-effect relations in time-limited psychotherapy for depression. *J Consult Clin Psychol*, 64(5), 927-935. doi: 10.1037/0022-006X.64.5.927
- Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der Helping Alliance Questionnaire (HAQ) von Luborsky: Möglichkeiten zur Evaluation des therapeutischen Prozesses von

- stationärer Psychotherapie. *Psychotherapeut*, 40, 23-32. doi: 10.1055/s-0030-1262820/id/30
- Bauer, S. (2004). *Ergebnismonitoring und Feedback: Mittel zur Optimierung stationärer Psychotherapie?* Dissertation. Fakultät für Sozial- und Verhaltenswissenschaften, Eberhard-Karls-Universität Tübingen.
- Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *J Pers Assess*, 82(1), 80-90. doi: 10.1207/s15327752jpa8201\_11
- Bauer, S., Okon, E., Meermann, R., & Kordy, H. (2012). Technology-enhanced maintenance of treatment gains in eating disorders: Efficacy of an intervention delivered via text messaging. *J Consult Clin Psychol*, 80(4), 700-706. doi: 10.1037/a0028030
- Baumann, U., Sodemann, U., & Tobien, H. (1980). Direkte versus indirekte Veränderungsdiagnostik. *Z Differ Diag Psychol*, 1, 201-216.
- Bergin, A.E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 217-270). New York, NY: Wiley.
- Berking, M., Orth, U., & Lutz, W. (2006). Wie effektiv sind systematische Rückmeldungen des Therapieverlaufs an den Therapeuten? Eine empirische Studie in einem stationär-verhaltenstherapeutischen Setting. *Z Klin Psychol Psychother*, 35, 21-29. doi: 10.1026/1616-3443.35.1.21
- Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *J Consult Clin Psychol*, 69(2), 197-204. doi: 10.1037/0022-006X.69.2.197
- Beutler, L. E. & Hamblin, D. L. (1986). Individualized outcome measures of internal change: Methodological considerations. *J Consult Clin Psychol*, 54(1), 48-53. doi: 10.1037/0022-006X.54.1.48
- Bickman, L., Douglas Kelley, S., Breda, C., De Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians mental health outcomes of youth: Results of a randomized trial. *Psychiatr Serv*, 62(12), 1423-1429. doi: 10.1176/appi.ps.002052011
- Boisvert, C. M., & Faust, D. (2003). Leading researchers' consensus on psychotherapy research findings: Implications for the teaching and conduct of psychotherapy. *Prof Psychol Res Pr*, 34(5), 508-513. doi: 10.1037/0735-7028.34.5.508
- Bohart, A. C., & Greaves Wade, A. (2013). The client in psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 219-257). Oxford, England: Wiley.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice*, 16(3), 252-260. doi: 10.1037/h0085885
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychother Res*, 25(1), 6-19. doi: 10.1080/10503307.2013.817696

- Brown, G. S., & Lambert, M. J. (1998, June). Tracking patient progress: Decision making for cases who are not benefiting from psychotherapy. Paper presented at the annual meeting of the Society for Psychotherapy Research, Snowbird, UT.
- Brown, G. S., & Jones, E. R. (2005). Implementation of a feedback system in a managed care environment: What are patients teaching us? *J Clin Psychol*, *61*(2), 187-98. doi: 10.1002/jclp.20110
- Carlier, I. V. E., Meuldijk, D., van Vliet, I. M., van Fenema, E., van der Wee, N. J. A., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract*, *18*(1), 104-110. doi: 10.1111/j.1365-2753.2010.01543.x
- Cecero, J. J., Fenton, L. R., Frankforter, T. L., Nich, C., & Carroll, K. M. (2001). Focus on therapeutic alliance: The psychometric properties of six measures across three instruments. *Psychother Theory Res Pract*, *38*, 1-11. doi: 10.1037/0033-3204.38.1.1
- Claiborn, C. D., & Goodyear, R. (2005). Feedback in psychotherapy. *J Clin Psychol*, *61*(2), 209-217. doi: 10.1002/jclp.20112
- Crawford, M. J., Thana, L., Farquharson, L., Palmer, L., Hancock, E., Bassett P., Clarke, J., & Perry, G. D. (2016). Patient experience of negative effects of psychological treatment: Results of a national survey. *Br J Psychiatry*, *208*(3): 260-265. doi: 10.1192/bjp.bp.114.162628
- Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J. L., Lambert, M. J., Gallop, R., McClure, B., Kulaga, A., & Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of Substance Abuse Treatment*, *42*(3), 301-309. doi: 10.1016/j.jsat.2011.09.003
- Cuipers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-analysis. *Clin Psychol Rev*, *32*(4), 280-291. doi: 10.1016/j.cpr.2012.01.003.
- Dawes, R. M. (1989). Experience and validity of clinical judgment: the illusory correlation. *Behavioral Sciences and the Law*, *7*(4), 457-467. doi: 10.1002/bsl.2370070404
- de Jong, K., Timman, R., Hakkaart-van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., & van Busschbach, J. (2014). The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychother Res*, *24*(6), 629-639. doi: 10.1080/10503307.2013.871079
- de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized controlled trial. *Psychother Res*, *22*(4), 464-474. doi: 10.1080/10503307.2012.673023
- de Jong-Meyer, R., Hautzinger, M., Rudolf, G., Strauß, W., & Frick, U. (1996). Prädiktions- und Verlaufsanalysen bei kombinierter psychologischer und medikamentöser Therapie bei endogen depressiven Patienten. *Z Klin Psychol*, *25*, 110-129.
- Delgadoillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: Development of a patient profiling and outcome feedback tool. *Behav Ther Res*, *79*, 15-22. doi: 10.1016/j.brat.2016.02.003

- Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M., Nevin, J., O'Hayon, H., Patel, U., Sainty, A. Spencer, P., & McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multi-site, open-label, cluster randomised controlled trial. *Lancet Psychiatry*, *5*, 564-572. doi: 10.1016/S2215-0366(18)30162-7
- Derogatis, L. R. (1986). *Symptom Checklist-90-Revised, Manual*. Towson, MD: Clinical Psychometric Research.
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognit Ther Res*, *14*, 469-482. doi:10.1007/BF01172968
- Dilcher, K., Mestel, R., Klingelhöfer, J., Köbel, W., Sprenger, B., & Stauss, K. (2000). Psychosomatische Kliniken. In N. Gerdes, H. Weidemann, & W. W. Jäckel (Eds.), *Die Protos-Studie* (pp. 173-202). Darmstadt: Steinkopff.
- Dilling, H., Mombour, W., Schmidt, M. H. (1993). *Internationale Klassifikation psychischer Störungen. ICD-10, Kapitel V (F)*. (2. Aufl). Huber, Bern
- Dinger, U., Strack, M., Leichsenring, F., Wilmers, F. & Schauenburg, H. (2008): Therapist effects on outcome and alliance in inpatient psychotherapy. *J Clin Psychol* *64*(3), 344-354. doi: 10.1002/jclp.20443
- Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education and workplace. *Psychological Science in the Public Interest*, *5*, 69-106. DOI: 10.1111/j.1529-1006.2004.00018.x
- Faller, H., Wagner, R. F., Vogel, H., Inselmann, U., Koepsell, K., Weiss, H., & Lang, H. (1997). Die Klientel einer psychotherapeutischen Poliklinik im Spiegel der Basisdokumentation. *Fortschr Neurol Psychiatr*, *65*(7), 313-322. doi: 10.1055/s-2007-996336
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *J Consult Clin Psychol*, *67*(4), 578-582. doi: 10.1037/0022-006X.67.4.578
- Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into early warning system. *Clin Psychol Psychother*, *8*, 231-242. doi: 10.1002/cpp.286
- Flückiger, C., Del Re, A. C., Wampold, B. E., Znoj., H., Caspar, F., & Jörg, U. (2012). Valuing clients' perspective and the effects on the therapeutic alliance: A randomized controlled study of an adjunctive instruction. *J Couns Psychol*, *59*(1), 18-26. doi: 10.1037/a0023648
- Franke, G. H. (1995). *Die Symptom-Checkliste von Derogatis, SCL-90-R – Deutsche Version*. Weinheim: Beltz.
- Franz, M., Janssen, P., Lensche, H., Schmidtke, V., Tetzlaff, M., Martin, K., Wöller, W., Hartkamp, N., Schneider, G., & Heuft, G. (2000). Effekte stationärer psychoanalytisch orientierter Psychotherapie – eine Multizenterstudie. *Z Psychosom Med Psychother*, *46*, 242-258.

- Freud S. (1912). Zur Dynamik der Übertragung. In A. Freud, M. Bonaparte, E. Bibring, W. Hoffer, E. Kris & O. Osakower (Eds., 1968), *Gesammelte Werke, Bd. VIII.* (pp. 364-374). Frankfurt a.M.: S. Fischer Verlag.
- Freud S. (1913). Zur Einleitung der Behandlung. . In A. Freud, M. Bonaparte, E. Bibring, W. Hoffer, E. Kris & O. Osakower (Eds., 1968), *Gesammelte Werke, Bd. VIII.* (pp. 454-478). Frankfurt a.M.: S. Fischer Verlag.
- Gallas, C., Puschner, B, Kühn, A., & Kordy, H. (2010). Dauer und Umfang ambulanter Psychotherapie und Implikationen für die Versorgungspraxis. *Psychother Psychosom Med Psychol*, 60(01), 5-13. doi: 10.1055/s-0029-1243258
- Gaston, L., Marmar, C. R. (1993). *Manual of the California Psychotherapy Alliance Scales*. Unpublished manuscript. University of California, San Francisco, CA: Department of Psychiatry.
- Gaston, L., Marmar, C.R., Gallagher, D. & Thompson, L.W. (1991). Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy processes. *Psychother Res*, 1(2), 104-112. doi: 10.1080/10503309112331335531
- Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Outcomes research in mental health – Systematic review. *Br J Psychiatry*, 181(1), 1-16. doi: 10.1192/bjp.181.1.8
- Grawe, K. (1982). *Implikationen und Anwendungsmöglichkeiten der Vertikalen Verhaltensanalyse für die Sichtweise und Behandlung psychischer Störungen.* (Forschungsberichte No. 5). Bern: Universität Bern.
- Grawe, K. (1992). Komplementäre Beziehungsgestaltung als Mittel zur Herstellung einer guten Therapiebeziehung. In J. Margraf & J. C. Brengelmann (Eds.), *Die Therapeut-Patient-Beziehung in der Verhaltenstherapie* (pp. 215-244). München: Röttger.
- Grawe, K., & Braun, U. (1994). Qualitätskontrolle in der Psychotherapiepraxis. *Z Klin Psychol*, 23, 242-267.
- Greenson, R. R. (1965). The working alliance and transference neurosis. *The Psychoanalytic Quarterly*, 34, 155-181.
- Grove, W. M. (2005). Clinical versus statistical prediction: the contribution of Paul E. Meehl. *J Clin Psychol*, 61(10), 1233-1243. doi: 10.1002/jclp.20179
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. *Psychology, Public Policy and Law*, 2(2), 293-323. doi: 10.1037/1076-8971.2.2.293
- Haas, E., Hill, R., Lambert, M. J., & Morrell, B. (2002). Do early responders to psychotherapy maintain treatment gains? *J Clin Psychol*, 58(9), 1157-1172. doi: 10.1002/jclp.10044
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *J Clin Psychol*, 61(2), 155-163. doi: 10.1002/jclp.20108
- Hansen, N.B., Lambert, M.J. & Forman, E.V. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clin Psychol*, 9(3), 329-343.

- Harmon, S.C., Hawkins, E.J., Lambert, M.J., Slade, K. & Whipple, J.S. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *J Clin Psychol*, *61*(2), 175-185. doi: 10.1002/jclp.20109
- Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychother Res*, *17*, 379-392. doi: 10.1080/10503300600702331
- Hatfield, D., McCullough, L., Frantz, S. H. B., Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clin Psychol Psychother*, *17*(1), 25–32. doi: 10.1002/cpp.656.
- Hawkins, E. J., Lambert, M. J., Vermeersch, D. A., Slade, K. L. & Tuttle, K. C. (2004). The therapeutic effects of providing patient progress information to therapists and patients. *Psychother Res*, *14*(3), 308-327. doi: 10.1093/ptr/kph027
- Hentschel, U. (2005). Die therapeutische Allianz. Teil 1: Die Entwicklungsgeschichte des Konzepts und moderne Forschungsansätze. *Psychotherapeut*, *50*(5), 305-317. doi: 10.1007/s00278-005-0440-3
- Heuft, G., Senf, W. (1998). Psy-BaDo-Basisdokumentation in der Psychotherapie. Deutsches Ärzteblatt, *95*, 2688-2692. A-2688 / B-2290 / C-2154
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V.S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *J Consult Clin Psychol*, *56*(6), 885-592.
- Horowitz, L. M., Strauß, B., & Kordy, H. (2000). *Inventar zur Erfassung interpersonaler Probleme – Deutsche Version* (2. überarbeitete und neunormierte Auflage). Weinheim: Beltz.
- Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37-69). New York: Oxford University Press.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *J Couns Psychol*, *36*(2), 223-233. doi: 10.1037/0022-0167.36.2.223
- Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy: A meta-analysis. *J Consult Clin Psychol*, *61*(4), 561-573. DOI: 10.1037//0022-006X.61.4.561
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *J Couns Psychol*, *38*(2), 139-149. Doi: 10.1037/0022-0167.38.2.139
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *Am Psychol*, *41*(2), 159-164. doi: 10.1037/0003-066X.41.2.159
- Howard, K. I., Moras, K., Brill, P., Martinovich, Z., & Lutz, W. (1996). The evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *Am Psychol*, *51*(10), 1059-1064. doi: 10.1037/2F0003-066X.51.10.1059

- Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behav Ther*, *15*(4), 336-352. doi: 10.1016/s005-7894(84)80002-7.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*, *59*(1), 12-19. doi: 10.1037/0022-006X.59.1.12
- Kadera, S. W., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough? A session-by-session analysis of the psychotherapy dose-effect relationship. *J Psychother Pract Res*, *5*(2), 1-20. <https://utswmed-ir.tdl.org/handle/2152.5/1277>
- Kastner, S., & Basler, H.-D. (1997). Messen Veränderungsfragebögen wirklich Veränderungen? *Schmerz*, *11*(4), 254-262. doi: 10.1007/s004820050094
- Kazdin, A.E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*, *63*(3), 146-159. doi: 10.1037/0003-066X.63.3.146
- Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., et al. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*, *7*. doi: 10.1002/14651858.CD011119.pub2.
- Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., et al. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *J Consult Clin Psychol*, *71*(6), 997-1006. doi: 10.1037/0022-006X.71.6.997
- Knaup, C., Kösters, M., Schöfer, D., Becker, T., & Puschner, B. (2009). Effect of feedback on treatment outcome in specialist mental healthcare: meta-analysis. *Brit J Psychiatry*, *195*(1), 15-22. doi: 10.1192/bjp.bp.108.053967
- Kopta, S.M., Howard, K.I., Lowry, J.L., & Beutler, L.E. (1994). Patterns of symptomatic recovery in psychotherapy. *J Consult Clin Psychol*, *62*(5), 1009-1016. doi: 10.1037/0022-006X.62.5.1009
- Kordy, H. (1992). Qualitätssicherung: Erläuterungen zu einem Reiz- und Modewort. *Z Psychosom Med*, *38*(4), 310-324. <https://www.jstor.org/stable/23996970>
- Kordy, H., & Gallas, C. (2007). Dokumentation und Qualitätssicherung in der Psychotherapie. In B. Strauß, F. Hohagen, & F. Caspar (Eds), *Lehrbuch Psychotherapie* (pp. 974-998). Göttingen: Hogrefe.
- Kordy, H., & Hannover, W. (1998). Beobachten, Dokumentieren, Bewerten, Steuern: Qualitätsmanagement in der stationären Psychotherapie. In A.-R. Laireiter & H. Vogel (Eds.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung* (pp. 355-374). Tübingen: dgvt-Verlag.
- Kordy, H., Hannover, W., & Richard, M. (2001). Computer assisted feedback-driven quality management for psychotherapy: the Stuttgart-Heidelberg model. *J Consult Clin Psychol*, *69*(2), 173-183. Doi: 10.1037/2F0022-006X.69.2.173
- Kordy, H. & Puschner, B. (2000). Aktive ergebnisorientierte Qualitätssicherung als Mittel zur Optimierung psychotherapeutischer Versorgung. In H. J. Freyberger, G. Heuft, & D. J.



- Ziegenhagen: *Ambulante Psychotherapie – Transparenz, Effizienz, Qualitätssicherung*. Stuttgart: Schattauer.
- Kordy, H., Wolf, M., Aulich, K., Bürgy, M., Hegerl, U., Hüsing, J., Puschner, B., Rummel-Kluge, C., Vedder, H., & Backenstrass, M. (2016). Internet-delivered disease management for recurrent depression: A multicenter randomized controlled trial. *Psychother Psychosom*, *85*(2), 91-8. doi: 10.1159/000441951.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*, *64*(3), 532-539. doi: 10.1037/0022-006X.64.3.532
- Lambert, M.J. (2013). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (Ed.). *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 169-218). Oxford, England: Wiley.
- Lambert, M. J. (2015). Progress feedback and the OQ-System: The past and the future. *Psychother Theor Res Pract Train*, *52*(4), 381-390. doi: 10.1037/pst0000027
- Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychother Theor Res Pract Train*, *38*(4), 357-361. doi: 10.1037/0033-3204.38.4.357
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clin Psychol Sci Pr*, *10*(3), 288-301. doi: 10.1093/clipsy.bpg025
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychother Res*, *11*(1), 49-68. doi: 10.1080/713663852
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., & Goates, M. K. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clin Psychol Psychoth*, *9*(2), 91-103. doi: 10.1002/cpp.324
- Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *J Clin Psychol*, *61*(2), 165-174. doi: 10.1002/jclp.20113
- Lambert, M.J., Hannover, W., Nisslmüller, K., Richard, M., & Kordy, H. (2002). Fragebogen zum Ergebnis von Psychotherapie: Zur Reliabilität und Validität der deutschen Übersetzung des Outcome Questionnaire 45.2 (OQ-45.2). *Z Klin Psychol Psychother*, *31*(1), 40-47. doi: 10.1026/0084-5345.31.1.40
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *J Consult Clin Psychol*, *69*(2), 159-172. Doi: 10.1037/2F0022-006X.69.2.159
- Lambert, M. J., Burlingame, G. M., Umphress, V. J., Hansen, N. B, Vermeersch, D. A., Clouse, G., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire.

- Clin Psychol Psychoth*, 3(4), 106–116. doi: 10.1002/(SICI)1099-0879(199612)3:4<249:AID-CPP106>3.0.CO;2-S
- Lambert, M. J., Kahler, M., Harmon, C., Burlingame, G. M., Shimokawa, K., & White, M.M. (2013). *Administration and Scoring Manual for the OQ-45.2 (Outcome Questionnaire)*. Orem, UT: American Professional Credentialing Services.
- Lambert, M.J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48(1), 72-79. doi: 10.1037/a0022238
- Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically derived and rationally derived methods for identifying patients at risk for treatment failure. *Clin Psychol Psychoth*, 9(3), 149-164. doi: 10.1002/cpp.333
- Lammers, C.-H. (2017). *Therapeutische Beziehung und Gesprächsführung*. Weinheim: Beltz.
- Liebherz S, Rabung S (2014). Do patients' symptoms and interpersonal problems improve in psychotherapeutic hospital treatment in Germany? - A systematic review and meta-analysis. *PLOS ONE* 9(8), doi: 10.1371/journal.pone.0105329
- Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy. A manual for supportive-expressive psychotherapy*. New York: Basic Books
- Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *J Consult Clin Psychol*, 69(2), 150-158. doi: 10.1037/2F0022-006X.69.2.150
- Lutz, W., Tholen, S., Schürch, E., & Berking, M. (2006). Die Entwicklung, Validität und Reliabilität von Kurzformen gängiger psychometrischer Instrumente zur Evaluation des therapeutischen Fortschrittes in Psychotherapie und Psychiatrie. *Diagnostica*, 52(1), 11-25. Doi: 10.1026/0012-1924.52.1.11
- Lutz, W., Stammer, H., Leeb, B., Dötsch, M., Bölle, M., & Kordy, H. (1996). Das Heidelberger Modell der Aktiven Internen Qualitätssicherung stationärer Psychotherapie. *Psychotherapeut*, 41(1), 25-35.
- Lutz, W., Martinovich, Z., & Howard, K.I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *J Consult Clin Psychol*, 67(4), 571-577. doi: 10.1037/0022-006X.67.4.571
- Lutz, W., Wittmann, W.W., Böhnke, J.R., Rubel, J., Steffanowski, A. (2012). Zu den Ergebnissen des Modellprojekts der Techniker-Krankenkasse zum Qualitätsmonitoring in der ambulanten Psychotherapie aus Sicht des wissenschaftlichen Evaluationsteams. *Psychother Psych Med*, 62(11), 413-417. Doi: 10.1055/s-0032-1327565
- Lutz, W., Wittmann, W. W., Böhnke, J .R., Rubel, J., & Steffanowski, A. (2013). Das Modellprojekt der TK zum Qualitätsmonitoring in der ambulanten Psychotherapie aus Sicht der Evaluatoren - Ein Plädoyer für mehr Psychotherapieforschung in Deutschland. *Psychother Psych Med*, 63(2), 25-228. doi: 10.1055/s-0033-1343399

- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *J Consult Clin Psychol*, *68*(3), 438-450. Doi: 10.1037/0022-006X.68.3.438
- McClintock, A. S., Perlman, M. R., McCarrick, S. M., Anderson, T., & Himawan, L. (2017). Enhancing psychotherapy process with common factors feedback: A randomized, clinical trial. *J Couns Psychol*, *64*(3), 247-260. doi: 10.1037/cou0000188
- McCullough Jr., J. P. (2006). *Treating Chronic Depression with Disciplined Personal Involvement: CBASP*. New York: Springer.
- McConaughy, E.A., Prochaska, J.O & Velicer, W.F. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychother Theor Res Pract Train*, *20*(3), 368-375. Doi: 10.1037/h0090198
- Meyer, F., & Schulte, D. (2002). Zur Validität der Beurteilung des Therapieerfolgs durch Therapeuten. *Z Kl Psych Psychoth*, *31*(1), 53-61. Doi: 10.1026/0084-5345.31.1.53
- Michalak, J., Kosfelder, J., Meyer, F., & Schulte, D. (2003). Messung des Therapieerfolgs: Veränderungsmaße oder retrospektive Erfolgsbeurteilung. *Z Kl Psych Psychoth*, *32*(2), 94-103. Doi: 10.1026//1616-3443.32.2.94
- Newnham, E. A., Hooke, G. R., & Page, A. C. (2010). Progress monitoring and feedback in psychiatric care reduces depressive symptoms. *J Affect Disord*, *127*(1-3), 139-146. Doi: 10.1016/j.jad.2010.05.003
- Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 134-166). New York: Wiley.
- Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clin Psychol Psychoth*, *10*(6), 361-373. Doi: 10.1002/cpp.383
- Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 307-389). New York: Wiley.
- Percevic, R. (2005). *Entwicklung und Evaluation eines computergestützten Testdarbietungssystems für therapiebegleitendes Ergebnismonitoring in der Psychotherapie*. Dissertation. Fakultät für Sozialwissenschaften, Universität Mannheim.
- Percevic, R., Bauer, S. & Kordy, H. (2004). Was ist klinisch bedeutsam an klinisch bedeutsamer Veränderung? Demonstration einer Validierungsstrategie am Beispiel des SCL-90R GSI. *Z Kl Psych Psychoth*, *33*(3), 165-171. Doi: 10.1026/1616-3443.33.3.165.
- Percevic, R., Gallas, C., Arikan, L., Mößner, M., & Kordy, H. (2006). Internet-gestützte Qualitätssicherung und Ergebnismonitoring in Psychotherapie, Psychiatrie und psychosomatischer Medizin. *Psychotherapeut*, *51*(5), 395-397. Doi: 10.1007/s00278-006-0504-z

- Percevic, R., Lambert, M. J., & Kordy, H. (2006). What is the predictive value of responses to psychotherapy for the future course? Empirical explorations and consequences for outcome monitoring. *Psychother Res*, 16(3), 364-373. doi: 10.1080/10503300500485524
- Percevic, R., Lambert, M.J. & Kordy, H. (2004). Computer supported monitoring of patient treatment response. *J Clin Psychol* 60(3), 285-300. Doi: 10.1002/jclp.10264
- Probst, T., Lambert, M. J., Dahlbender, R. W., Loew, T. H., & Tritt, K. (2014). Providing patient progress feedback and clinical support tools to therapists: Is the therapeutic process of patients on-track to recovery enhanced in psychosomatic in-patient therapy under the conditions of routine practice? *J Psychosom Res*, 76(6), 477-84. doi: 10.1016/j.jpsychores.2014.03.010.
- Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *J Psychosom Res*, 75(3), 255-61. doi: 10.1016/j.jpsychores.2013.07.003.
- Probst, T., v. Heymann, F., Zaudig, M., Konermann, J., Lahmann, C., Loew, T., & Tritt, K. (2009): Effektivität stationärer psychosomatischer Krankenhausbehandlung - Ergebnisse einer multizentrischen Katamnesestudie. *Z Psychosom Med Psychother* 55(4), 409-420. <https://www.jstor.org/stable/23872285>
- Puschner, B., Schöfer, D., Knaup, C., & Becker, T. (2009). Outcome management in in-patient psychiatric care. *Acta Psychiatr Scand*, 120(4), 308-319. Doi: 10.1111/j.1600-0447.2009.01397.x
- Puschner, B., Wolf, M., & Kraft, S. (2009). Helping alliance and outcome in psychotherapy: What predicts what in routine outpatient treatment? *Psychother Res*, 18(2), 167-178. Doi: 10.1080/10503300701367984
- Puschner, B., Kraft, S., Kächele, H., & Kordy, H. (2007). Course of improvement over 2 years in psychoanalytic and psychodynamic outpatient psychotherapy. *Psychol Psychother*, 80(1), 51-68. Doi: 10.1348/147608306X107593
- Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.). *Social psychology and evaluation*, New York, NY: Guildford Press.
- Rudolf, G., Manz, R. (1993). Zur prognostischen Bedeutung der therapeutischen Arbeitsbeziehung aus der Perspektive von Patienten und Therapeuten. *Psychother Psychosom Med Psychol* 43, 193-199-
- Safran, J. D., Muran, J. C., Samstag, L. W., & Stevens, C. (2002). Repairing alliance ruptures. In J. C. Norcross (Ed.), *Psychotherapy relationships that work*. New York: Oxford University Press.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research and practice. *J Clin Psychol*, 61(2):145-153. Doi: 10.1002/jclp.20107
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *J Consult Clin Psychol*, 80(4), 535-46. doi: 10.1037/a0028898.

- Schauenburg, H., Dinger, U., & Strack, M. (2005). Zur Bedeutung des Einzeltherapeuten für das Therapieergebnis in der stationären Psychotherapie - eine Pilotstudie. *Psychother Psych Med*, 55(7), 339-346. DOI: 10.1055/s-2004-834750
- Scheidt, C. E., Seidenglanz, K., Dieterle, W., Hartmann, A., Bowe, N., Hillenbrand, D., Scudlek, G., Strasser, F., Strasser, P., & Wirsching M. (1999). Basisdaten zur Qualitätssicherung in der ambulanten Psychotherapie. Ergebnisse einer Untersuchung in 40 psychotherapeutischen Fachpraxen. Teil 2: Behandlungsverlauf und Behandlungsergebnisse. *Psychotherapeut*, 44(2), 83-93.
- Scheidt, C. E., Brockmann, J., Caspar, F., Rudolf, G., Stangier, U., Vogel, H. (2012). Das Modellprojekt der Techniker-Krankenkasse: Eine Kommentierung der Ergebnisse aus der Sicht des wissenschaftlichen Projektbeirates. *Psychother Psych Med*, 62(11), 405-412. DOI: 10.1055/s-0032-1321781
- Schepank, H. (1995). *Der Beeinträchtigungs-Schwere-Score (BSS)*. Göttingen: Beltz Test
- Schmidt, J., Lamprecht, F., Wittmann, W.W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *Psychother Psychosom Med*, 39(7): 248-255.
- Schulte, D. (1993). Wie soll Therapieerfolg gemessen werden? *Z Klin Psychol*, 22, 374-393.
- Schulte, D., & Meyer, F. (2002). Woran orientieren sich Therapeuten bei ihrer Sitzungsbeurteilung und Erfolgsprognose? *Z Kl Psych Psychoth*, 31(4), 257-265. Doi: 10.1026/0084-5345.31.4.257
- Scott, J., & Young, A .H. (2016). Psychotherapies should be assessed for both benefit and harm. *Br J Psychiatry*, 208(3), 208-209. doi: 10.1192/bjp.bp.115.169060.
- Selbmann, H. K. (1996). Zum Stand der Qualitätssicherung in der ambulanten Versorgung in Deutschland - Ergebnisse einer analytischen Bestandsaufnahme. In Bundesministerium für Gesundheit (Ed.). *Qualitätssicherung in der ambulanten Versorgung*. Baden-Baden: Nomos
- Shapiro, D. A., & Shapiro D. (1982). Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychol Bull*, 92(3), 581-604. Doi: 10.1037/0033-2909.92.3.581
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-Analytic and mega-analytic review of a psychotherapeutic quality assurance system. *J Consult Clin Psychol*, 78(3), 298-311. Doi: 10.1037/a0019247
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk for treatment failure. *Psychother Res*, 22, 638-647. doi: 10.1080/10503307
- Simon, W., Lambert, M.J., Busath, G, Vazquez, A., Berkeljon, A., Hyer, K., Granley, M., Berrett, M. (2013). Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomized controlled study. *Psychother Res*, 23, 287-300. doi: 10.1080/10503307
- Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clin Psychol Psychother*, 15(5), 287-303. doi: 10.1002/cpp.594.

- Smith M. L., Glass G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: John Hopkins University Press.
- Sperry, L., Brill, P. L., Howard, K. I., & Grissom, G. R. (1996). *Treatment outcomes in psychotherapy and psychiatric interventions*. New York: Brunner/Mazel.
- Steffanowski, A., Löschmann, C. Schmidt, J., Wittmann, W. W., Nübling, R. (2007). *Meta-Analyse der Effekte stationärer psychosomatischer Rehabilitation: Mesta-Studie*. Bern: Huber.
- Steffanowski, A., Fembacher, A., Kramer, D., Pfaffinger, I., Kriz, D., Glahn, E.M., Wittmann, W.W. (2012). Praxisübergreifende Dokumentation der Ergebnisqualität ambulanter Psychotherapie in Bayern: Ergebnisse der Pilotstudie "QS-PSY-BAY". *Z Kl Psych Psychoth*, 30(1), 10-21. doi: 10.1026/1616-3443/a000124
- Sterba, R. (1934). The fate of the ego in analytic therapy. *International Journal of Psychoanalysis*, 15, 117-126.
- Stieglitz, R.-D., Hiller, W. (2014). Strategien und Instrumente der Veränderungsmessung. *ZPPP*, 62(2), 101-111. doi: 10.1024/1661-4747/a000187
- Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., Iveson, M., & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *J Consult Clin Psychol*, 71(1), 14-21. DOI: 10.1037//0022-006X.71.1.14
- Strauß, B., Linden, M., Haupt, M.-L., & Kaczmarek, S. (2012). Systematik und Häufigkeit von unerwünschten Wirkungen, Nebenwirkungen und Fehlentwicklungen in der Psychotherapie. *Psychotherapeut*, 57(5), 385-394. doi: 10.1007/s00278-012-0932-x
- Strupp, H. H., Horowitz, L. M. & Lambert, M. J (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery*. Washington, DC: American Psychological Association.
- Swift, J. K., Callahan, J. L., Rousmaniere, T. G., Whipple, J. L., Dexter, K., & Wrape, E. R. (2015). Using client outcome monitoring as a tool for supervision. *Psychother Theory Res Pract*, 52(2), 180-184. DOI: 10.1037/a0037659
- Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *J Consult Clin Psychol*, 67, 894-904.
- Thomä, H., & Kächele, H. (2006). *Psychoanalytische Therapie. Grundlagen*. 3., überarb. u. aktualisierte Aufl. Heidelberg: Springer.
- Tritt, K., Heymann, F. von, Loew, T., Benker, B., Bleichner, F., Buchmüller, R., Findeisen, P., Galuska, J., Kalleder, W. Lettner, F., Michelitsch, B., Stadtmüller, G., & Zaudig, M. (2003): Patienten in stationärer psychosomatischer Krankenhausbehandlung: Patientencharakterisierung und Behandlungsergebnisse anhand der Psy-BaDo-PTM. *Psychother Psychiatr Psychother Med Klin Psychol* 8, 244-251.
- Vogel, H., Brockmann, J., Caspar, F., Rudolf, G., Stangier, U., Scheidt, C.E. (2013). Zu den Ergebnissen des Modellprojektes der Techniker-Krankenkasse zum Qualitätsmonitoring in der ambulanten Psychotherapie - eine Replik des wissenschaftlichen Beirats. *Psychother Psych Med*, 63(6), 229-231. Doi: 10.1055/s-0033-1341501

- von Zerssen, D. (1986 a). Beschwerden-Liste BL. In Collegium Internationale Psychiatriae Scalarum (Eds.), *Internationale Skalen für Psychiatrie*. Weinheim: Beltz.
- von Zerssen, D. (1986 b). Depressivitäts-Skala DS. In Collegium Internationale Psychiatriae Scalarum (Eds.), *Internationale Skalen für Psychiatrie*. Weinheim: Beltz.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychol Rep*, *110*(2), 639-644. doi: 10.2466/02.07.17.PR0.110.2.639-644
- Wampold, B. E. & Brown, J. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *J Consult Clin Psychol*, *73*(5), 914–23. Doi: 10.1037/0022-006X.73.5.914
- Wampold, B.E., & Imel, Z.E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2<sup>nd</sup> ed.). New York, NY: Routledge.
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *J Couns Psychol* *50*(1), 59-68. Doi: 10.1037/2F0022-0167.50.1.59
- World Health Organization. Division of Mental Health (1997). *Quality assurance in mental health*. Geneva: World Health Organization.
- Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy. A practitioner's guide*. New York, NY: Guildford Press.
- Zetzel, E. R. (1956). Current concepts of transference. *Journal of Psychoanalysis*, *37*, 369-376.
- Zielke, M., & Kopf-Mehnert, C. (1978). *Veränderungsfragebogen des Erlebens und Verhaltens*. Weinheim: Beltz.
- Zimmer, B., & Moessner, M. (2012). Therapieevaluation in der stationären Psychotherapie mit Web-AKQUASI. *PiD*, *13*(1), 68-71. doi: 10.1055/s-0031-1298936
- Znoj, H. (2005). Die therapeutische Beziehung aus verhaltenstherapeutischer Sicht. In W. Rössler (Ed.), *Die therapeutische Beziehung* (pp. 81-95), Heidelberg: Springer.

## 7 TABELLARISCHER ANHANG

- Anhang A.** Kennwerte für die Berechnung von reliabler und klinisch signifikanter Veränderung für die HAQ-Skala „Beziehungszufriedenheit“
- Anhang B.** Prozentsatz an reliablen Veränderungen auf den SCL-90-R-Subskalen zwischen Aufnahme und Entlassung für die drei Untersuchungsgruppen und die Gesamtstichprobe
- Anhang C.** Prozentsatz an reliablen Veränderungen auf den IIP-Subskalen zwischen Aufnahme und Entlassung für die drei Untersuchungsgruppen und die Gesamtstichprobe



**Anhang A.** Kennwerte für die Berechnung von reliabler und klinisch signifikanter Veränderung für die HAQ-Skala „Beziehungszufriedenheit“

Kennwert	Ergebnis für die klinikinterne Normstichprobe (n=348)
sd	1,03
$\alpha$	0,89
$S_e$	0,346
$S_{diff}$	0,489
68. Perzentil	1,33
84. Perzentil	0,67

**Anhang B.** Prozentsatz an reliablen Veränderungen auf den SCL-90-R-Subskalen zwischen Aufnahme und Entlassung für die drei Untersuchungsgruppen und die Gesamtstichprobe

	Gruppe 1			Gruppe 2			Gruppe 3			Gesamt		
	(n=163)			(n=168)			(n=206)			(n=537)		
	RC +	Kein RC	RC -	RC +	Kein RC	RC -	RC +	Kein RC	RC -	RC +	Kein RC	RC -
SCL 1	27,0	66,3	6,7	33,9	58,9	7,1	31,6	59,2	9,2	30,9	61,3	7,8
SCL 2	53,4	41,7	4,9	53,0	44,6	2,4	48,1	49,5	2,4	51,2	45,6	3,2
SCL 3	53,4	38,7	8,0	60,7	34,5	4,8	51,9	45,1	2,9	55,1	39,9	5,0
SCL 4	56,4	34,4	9,2	67,3	26,8	6,0	63,1	31,6	5,3	62,4	30,9	6,7
SCL 5	46,0	44,2	9,8	51,8	38,1	10,1	41,7	51,0	7,3	46,2	44,9	8,9
SCL 6	33,7	54,0	12,3	31,0	57,1	11,9	34,1	60,0	5,9	33,0	57,3	9,7
SCL 7	38,0	56,4	5,5	38,7	55,4	6,0	35,1	59,0	5,9	37,1	57,1	5,8
SCL 8	39,3	51,5	9,2	42,9	53,6	3,6	40,8	54,9	4,4	41,0	53,4	5,6
SCL 9	41,7	49,1	9,2	47,6	44,0	8,3	43,7	49,5	6,8	44,3	47,7	8,0
GSI	65,0	24,5	10,4	72,0	15,5	12,5	72,3	21,4	6,3	70,0	20,5	9,5

*Anmerkungen:* RC +: reliable Verbesserung, kein RC: keine Veränderung, RC -: reliable Verschlechterung, SCL 1: SCL-90-R Skala 1 (Somatisierung), SCL 2: SCL-90-R Skala 2 (Zwanghaftigkeit), SCL 3: SCL-90-R Skala 3 (Unsicherheit im Sozialkontakt), SCL 4: SCL-90-R Skala 4 (Depressivität), SCL 5: SCL-90-R Skala 5 (Ängstlichkeit), SCL 6: SCL-90-R Skala 6 (Aggressivität/ Feindseligkeit), SCL 7: SCL-90-R Skala 7 (Phobische Angst), SCL 8: SCL-90-R Skala 8 (Paranoides Denken), SCL 9: SCL-90-R Skala 9 (Psychotizismus).

**Anhang C.** Prozentsatz an reliablen Veränderungen auf den IIP-Subskalen zwischen Aufnahme und Entlassung für die drei Untersuchungsgruppen und die Gesamtstichprobe

	Gruppe 1			Gruppe 2			Gruppe 3			Gesamt		
	(n=163)			(n=168)			(n=206)			(n=537)		
	RC +	Kein RC	RC -	RC +	Kein RC	RC -	RC +	Kein RC	RC -	RC +	Kein RC	RC -
<i>IIP 1</i>	7,4	87,7	4,9	7,9	86,0	6,1	8,9	87,1	4,0	8,1	86,9	4,9
<i>IIP 2</i>	7,4	87,7	4,9	11,6	84,8	3,7	7,4	91,1	1,5	8,7	88,1	3,2
<i>IIP 3</i>	7,4	90,1	2,5	12,2	83,5	4,3	11,4	88,1	0,5	10,4	87,3	2,3
<i>IIP 4</i>	14,8	83,3	1,9	19,5	79,3	1,2	13,9	85,1	1,0	15,9	82,8	1,3
<i>IIP 5</i>	20,4	75,9	3,7	14,6	79,9	5,5	13,4	83,7	3,0	15,9	80,1	4,0
<i>IIP 6</i>	11,7	86,4	1,9	17,7	80,5	1,8	13,9	82,2	4,0	14,4	83,0	2,7
<i>IIP 7</i>	8,0	90,1	1,9	14,0	82,9	3,0	13,9	82,2	4,0	12,1	84,8	3,0
<i>IIP 8</i>	8,6	86,4	4,9	8,5	87,2	4,3	8,4	89,1	2,5	8,5	87,7	3,8

*Anmerkungen:* RC +: reliable Verbesserung, kein RC: keine Veränderung, RC -: reliable Verschlechterung, IIP 1: SCL 2: SCL-90-R Skala 2 (Zwanghaftigkeit), SCL 3: SCL-90-R Skala 3 (Unsicherheit im Sozialkontakt), SCL 4: SCL-90-R Skala 4 (Depressivität), SCL 5: SCL-90-R Skala 5 (Ängstlichkeit), SCL 6: SCL-90-R Skala 6 (Aggressivität/ Feindseligkeit), SCL 7: SCL-90-R Skala 7 (Phobische Angst), SCL 8: SCL-90-R Skala 8 (Paranoides Denken), SCL 9: SCL-90-R Skala 9 (Psychotizismus).

## 8 LEBENS LAUF

### PERSONALIEN

Name und Vorname: Christine Gallas  
Geburtsdatum: 04.10.1972 in München  
Familienstand: verheiratet

### SCHULISCHER WERDEGANG

1979 – 1992                      Grundschule und Gymnasium in München  
Juli 1992                        Abitur (Durchschnittsnote 1,9), Edith-Stein-Gymnasium München

### UNIVERSITÄRER WERDEGANG

WS 1992 – WS 1999            Psychologie-Studium an der Universität Konstanz und der  
University of Ulster at Coleraine (GB)  
Diplomarbeit: „Hunger nach Erfolg: Eine empirische Studie zum  
Essverhalten bei begabten Jugendlichen“, Note: sehr gut  
22.03.2000                      Diplom, Gesamtnote: sehr gut

### BERUFLICHER WERDEGANG

Seit 04/2016                    Psychologische Psychotherapeutin, Hochschulambulanz Klinische  
Psychologie, Zentralinstitut für Seelische Gesundheit Mannheim  
07/2014 – 03/2016            Elternzeit  
04/2013 – 06/2014            Psychotherapeutin in Ausbildung in der Hochschulambulanz  
Klinische Psychologie, Zentralinstitut für Seelische Gesundheit  
11/2011 – 03/2013            Elternzeit  
05/2007 – 10/2011            Wissenschaftliche Mitarbeiterin am Zentralinstitut für Seelische  
Gesundheit  
04/2002 – 03/2007            Wissenschaftliche Mitarbeiterin an der Forschungsstelle für  
Psychotherapie, Universitätsklinikum Heidelberg  
01/2001 – 03/2002            Beraterin, Plocek Unternehmensberatung, München  
04/2000 – 12/2000            Angestellte bei custo med GmbH, Ottobrunn

## PUBLIKATIONEN

- Dreßing, H., Bailer, J., Anders, A., Wagner, H., & Gallas, C. (2014). Cyberstalking in a large sample of social network users: Prevalence, characteristics, and impact upon victims. *Cyberpsychol Behav Soc Netw.*, *17*(2), 61-67. doi: 10.1089/cyber.2012.0231.
- Salize, H.J., Jacke, C., Gallas, C., & Stamm, K. (2013). Optimized quality of care for affective disorders by health insurance-based case-management: a controlled cost-study. *Psychiatr Prax*, *41*(8), 432-438. doi: 10.1055/s-0033-1349557.
- Dreßing, H., Anders, A., Gallas, C., Bailer, J. (2011). Cyberstalking: Prävalenz und Auswirkungen auf die Betroffenen. *Psychiatr Prax*, *38*(7), 336-341. doi: 10.1055/s-0031-1276860.
- Moessner, M., Gallas, C., Haug, S., Kordy, H. (2011). The Clinical Psychological Diagnostic System (KPD-38): Sensitivity to change and validity of a self-report instrument for outcome monitoring and quality assurance. *Clin Psycholog Psychother*, *18*(4), 331-338. doi: 10.1002/cpp.717.
- Gallas, C., Puschner, B., Kühn, A., Kordy, H. (2010). Dauer und Umfang ambulanter Psychotherapie und Implikationen für die Versorgungspraxis. *Psychother Psychosom Med Psychol*, *60* (1), 5-13. doi: 10.1055/s-0029-1243258.
- Dreßing, H., Klein, U., Bailer, J., Gass, P., Gallas, C. (2009). Cyberstalking. *Nervenarzt*, *80*(7), 833-836. doi: 10.1007/s00115-009-2732-2.
- Gallas, C., Bindeballe, N., Gass, P., Dreßing H. (2009). Implementierung eines strukturierten therapeutischen Gruppenprogramms für Stalking-Opfer: ein Pilotprojekt. *Psychotherapeut*, *54*(3), 199-204.
- Mößner, M., Zimmer, B. Gallas, C., Percevic, R., Kordy, H. (2009). Das Klinisch psychologische Diagnosesystem 38 (KPD-38): Validierung an einer multizentrischen Patientenstichprobe. *Klinische Diagnostik und Evaluation* (2), 194-204.
- Gallas, C., Kächele, H., Kraft, S., Kordy, H., Puschner, B. (2008). Inanspruchnahme, Verlauf und Ergebnis ambulanter Psychotherapie: Befunde der TRANS-OP Studie und deren Implikationen für die Richtlinienpsychotherapie. *Psychotherapeut* *53*(6), 414-423.
- Dreßing, H., Gallas, C., Bindeballe N., Gass P. (2008). Arbeits- und Sozialmedizinische Aspekte von Stalking. *Arbeitsmedizin, Sozialmedizin, Umweltmedizin*, *43*(6), 314-318.
- Dreßing, H., Bindeballe, N., Gallas, C., Gass, P. (2008). Stalking: Klinische und Forensische Bedeutung. *Journal für Neurologie, Neurochirurgie und Psychiatrie* *9*(4), 20-27.
- Dreßing, H., Bindeballe, N., Gallas, C., Gass, P. (2008). Aufgaben der Allgemeinpsychiatrie und der Forensischen Psychiatrie bei Stalking. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, *56*(2), 111-119.
- Haug, S., Strauß, B., Gallas, C., Kordy, H. (2008). New prospects for process research in group therapy: text-based process variables in psychotherapeutic Internet-chat groups. *Psychother Res*, *18*(1), 88-96. doi: 10.1080/10503300701368008.

- Percevic, R., Gallas, C., Arikan, L., Mößner, Kordy, H. (2006). Internet-gestützte Qualitätssicherung und Ergebnismonitoring in Psychotherapie, Psychiatrie und psychosomatischer Medizin. *Psychotherapeut*, 51, 395-397.
- Percevic, R., Gallas, C., Wolf, M., Haug, S., Hünerfauth, T., Schwarz, M., Kordy, H. (2005). Das Klinisch Psychologische Diagnosesystem (KPD-38): Entwicklung, Normierung und Validierung eines Selbstbeurteilungsbogen für den Einsatz in Qualitätssicherung und Ergebnismonitoring in der Psychotherapie und psychosomatischen Medizin. *Diagnostica*, 51, 134-144.
- Wolf, M., Gallas, C., & Kordy, H. (2005). Zur diskriminativen Validität des Inventars zur Erfassung Interpersonaler Probleme (IIP): Interpersonale Probleme von Patienten mit Persönlichkeitsstörungen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 34(4), 233-240.

### **Buchbeiträge**

- Gallas, C., Klein, U., & Dreßing, H. (2010). *Beratung und Therapie von Stalking-Opfern – Ein Leitfaden für die Praxis*. Bern: Hans Huber.
- Kordy, H., Gallas, C., & Zimmer, B. (2011). Grundlagen - Dokumentation und Qualitätssicherung. In W. Senf, M. Broda (Eds.) *Praxis der Psychotherapie. Ein integratives Lehrbuch*. (5. Auflage). Stuttgart: Thieme.
- Gallas, C. (2010). Beratung und psychotherapeutische Unterstützung von Stalkingopfern. In Weisser Ring e.V. (Ed.). *Stalking: Wissenschaft, Gesetzgebung und Opferhilfe*. (pp. 54-70). Baden-Baden: Nomos Verlag
- Kordy, H., & Gallas, C. (2007). Dokumentation und Qualitätssicherung in der Psychotherapie. In B. Strauß, F. Hohagen, & F. Caspar (Eds.), *Lehrbuch Psychotherapie* (pp. 974-998). Göttingen: Hogrefe.

## 9 DANKSAGUNG

Von Herzen danke ich allen Personen, die an dieser Arbeit beteiligt waren.

An erster Stelle gilt mein Dank meiner Doktormutter Prof. Dr. Christine Kühner für ihre fachliche und emotionale Unterstützung und ihre Ermutigung, das Ziel auch nach langen Pausen nicht aus den Augen zu verlieren.

Prof. Dr. Peter Kirsch und Priv.-Doz. Dr. Anne Dyer danke ich herzlich für ihre Bereitschaft, die Begutachtung dieser Arbeit zu übernehmen.

Dr. Hans Kordy, der während der Planung und Umsetzung der Studie Leiter der Forschungsstelle für Psychotherapie am Universitätsklinikum Heidelberg war, danke ich für die gemeinsame Entwicklung des Themas dieser Arbeit.

Für die Möglichkeit, die Studie an der Klinik Tiefenbrunn durchzuführen, sowie wertvolle konzeptionelle und organisatorische Anregungen, danke ich dem damaligen Leiter der Abteilung Forschung und Dokumentation, Prof. Dr. Falk Leichsenring. Ein besonderer Dank gilt Herrn Dipl.-Psych. Ulrich Jäger, der die Datenerhebung vor Ort plante und organisierte. Ohne sein großes Engagement wäre die Untersuchung in dieser Form nicht möglich gewesen. Außerdem danke ich den beteiligten Therapeutinnen und Therapeuten sowie den Patientinnen und Patienten für ihre Bereitschaft, sich auf ein neues Projekt einzulassen, dafür Zeit zu investieren und Informationen über sich preiszugeben.

Ein großer Dank gilt Frau Dipl.-Math. Iris Reinhard, die mit großer Kompetenz und Geduld Fragen zur Datenauswertung diskutierte und klärte.

An letzter Stelle genannt, obwohl die wichtigste einnehmend, danke ich Steffen Gallas, der diesen langen Weg mit mir gegangen ist. Ohne ihn hätte ich diese Arbeit nicht fertig stellen können.

Mannheim, im Juli 2019