Der Medizinischen Fakultät der Georg-August-Universität
eingereicht von Prof. Dr. med. C. Spitzer

Nonresponse und Misserfolge in der stationären Psychotherapie

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von

Ann-Kathrin Albers

aus

Mettingen

Göttingen 2020

Die vorliegende Dissertation wurde in Rosdorf im ASKLEPIOS Fachklinikum Tiefenbrunn im Zeitraum von 2015 bis 2020 unter der Betreuung von Prof. Dr. med. C. Spitzer angefertigt.

Dekan:	Prof. Dr. med. W. Brück
Referent/in	Prof. Dr. med. C. Spitzer
Ko-Referent/in:	Prof. Dr. med. K. Schnell
Drittreferent/in:	Prof. Dr. mult. T. Meyer

Datum der mündlichen Prüfung: 18.11.2021

Hiermit erkläre ich, die Dissertation mit dem Titel „Non-response und Misserfolge in der stationären Psychotherapie“ eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den …… ……

(Unterschrift)

Inhaltsverzeichnis

Abbildungsverzeichnis ..... **III**

Tabellenverzeichnis ..... **IV**

Abkürzungsverzeichnis ..... **VI**

1 Einleitung **1**

1.1 Stationäre Psychotherapie 2

1.2 Misserfolge in der Psychotherapie 4

1.2.1 Abgrenzungen der Begrifflichkeiten in der Misserfolgsforschung 5

1.2.2 Häufigkeit von Nonresponse und Verschlechterung 6

1.2.3 Korrelate von Nonresponse und Verschlechterung 6

1.2.4 Persönlichkeitsstörungen 12

1.3 Veränderungsmessung 14

1.3.1 Direkte Veränderungsmessung 15

1.3.2 Indirekte Veränderungsmessung 17

1.3.3 Direkte und indirekte Veränderungsmessung im Vergleich 19

1.3.4 Selbst- und Fremdbeurteilungsverfahren 19

1.3.5 Probleme und Ansprüche bei der Veränderungsmessung 20

1.4 Multimodale Diagnostik 22

1.5 Zusammenfassung und eigene Fragestellung 23

1.5.1 Zusammenfassung und Bewertung der aktuellen Forschungslage 23

1.5.2 Fragestellungen und Ziele dieser Arbeit 24

2 Patienten, Material und Methoden **25**

2.1 Ablauf der Studie 25

2.2 Ein- und Ausschlusskriterien 25

2.3 Untersuchungsinstrumente 26

2.3.1 Die Symptom-Checkliste SCL-90-R von L. R. Derogatis 28

2.3.2 Inventar zur Erfassung Interpersonaler Probleme IIP 29

2.3.3 Der Beeinträchtigungs-Schwere-Score BSS 32

2.3.4 Basisdokumentation und Fragen zur direkten Veränderungsmessung 34

2.4 Datenverarbeitung und statistische Verfahren 35

2.4.1 Methoden zur Stichprobenbeschreibung und Prüfung der Repräsentativität 35

2.4.2 Methoden zur Untersuchung der ersten Forschungsfrage 35

2.4.3 Methoden zur Untersuchung der zweiten und dritten Forschungsfrage 36

2.4.4 Methoden zur Untersuchung soziodemographischer Aspekte 38

2.4.4.1 Prüfung auf das Vorliegen von soziodemographischen Unterschieden zwischen den analysierten Gruppen 38

2.4.4.2 Prüfung des Einflusses soziodemographischer Parameter auf das Therapieergebnis 39

3 Ergebnisse **43**

3.1 Beschreibung der Stichprobe 43

3.1.1 Soziodemographische Merkmale 43

3.1.2 Klinische Charakteristika ... 45
3.1.3 Vergleich zwischen der untersuchten Stichprobe und der Stichprobe vor Anwendung der Einschlusskriterien ... 48
3.2 Häufigkeit von Nonresponse und Verschlechterung ... 51
3.2.1 Direkte Veränderungsmessung ... 51
3.2.2 Indirekte Veränderungsmessung ... 52
3.3 Nonresponse und Verschlechterung im Geschlechtervergleich ... 55
3.3.1 Direkte Veränderungsmessung ... 55
3.3.2 Indirekte Veränderungsmessung ... 57
3.4 Nonresponse und Verschlechterungen bei Persönlichkeitsstörungen ... 61
3.4.1 Ergebnisse für Personen mit und ohne Persönlichkeitsstörungen ... 61
3.4.1.1 Direkte Veränderungsmessung ... 61
3.4.1.2 Indirekte Veränderungsmessung ... 63
3.4.2 Ergebnisse für Personen mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen ... 67
3.4.2.1 Direkte Veränderungsmessung ... 67
3.4.2.2 Indirekte Veränderungsmessung ... 69
3.5 Soziodemographische Unterschiede der untersuchten Patientengruppen ... 73

4 Diskussion ... 79

4.1 Zusammenfassung der Studienergebnisse ... 79
4.2 Diskussion der Methoden ... 80
4.2.1 Studiendesign ... 80
4.2.2 Ein- und Ausschlusskriterien ... 82
4.2.3 Untersuchungsinstrumente ... 85
4.3 Diskussion der Ergebnisse ... 87
4.3.1 Repräsentativität der Stichprobe ... 87
4.3.1.1 Soziodemographische Aspekte ... 87
4.3.1.2 Klinische Merkmale ... 88
4.3.1.3 Vergleich der Stichproben vor und nach Anwendung der Einschlusskriterien ... 93
4.3.2 Diskussion der 1. Forschungsfrage ... 94
4.3.3 Diskussion der 2. Forschungsfrage ... 105
4.3.4 Diskussion der 3. Forschungsfrage ... 107
4.3.5 Soziodemographische Parameter als Confounder ... 109
4.4 Limitationen der Untersuchung ... 111
4.5 Schlussfolgerungen und Ausblick ... 112

5 Zusammenfassung ... 114

6 Literaturverzeichnis ... 116

Abbildungsverzeichnis

Abbildung 1: Verteilung der Gesunden und Kranken 18
Abbildung 2: Circumplexskalen des IIP 30
Abbildung 3: Zusammenhänge zwischen den Parametern der logistischen Regression 40
Abbildung 4: Beurteilung des Behandlungserfolgs zum Entlassungszeitpunkt anhand der Veränderung der Symptomatik aus Patienten- und Therapeutenperspektive 52
Abbildung 5: Behandlungsergebnisse (Erfolg, Nonresponse, Verschlechterung) im Rahmen der indirekten Veränderungsmessung, ohne Darstellung der Pseudogesunden 53
Abbildung 6: Behandlungsergebnisse im Rahmen der indirekten Veränderungsmessung 53
Abbildung 7: Selbsteinschätzung der Veränderung der Symptomatik differenziert nach dem Geschlecht 56
Abbildung 8: Therapeuteneinschätzung der Veränderung der Symptomatik differenziert nach dem Geschlecht 56
Abbildung 9: Summenscores differenziert nach dem Geschlecht und Messzeitpunkt 60
Abbildung 10: Veränderung der Symptomatik aus Patientenperspektive differenziert nach dem Vorliegen einer Persönlichkeitsstörung 62
Abbildung 11: Veränderung der Symptomatik aus Therapeutenperspektive differenziert nach dem Vorliegen einer Persönlichkeitsstörung 62
Abbildung 12: Mittelwerte der Summenscores differenziert nach dem Vorliegen von Persönlichkeitsstörungen und dem Messzeitpunkt 66
Abbildung 13: Veränderung der Symptomatik aus Patientenperspektive differenziert nach Persönlichkeitsstörungsclustern 68
Abbildung 14: Veränderung der Symptomatik aus Therapeutenperspektive differenziert nach Persönlichkeitsstörungsclustern 68
Abbildung 15: Summenscores differenziert nach dem Persönlichkeitsstörungscluster und dem Messzeitpunkt 72

Tabellenverzeichnis

Tabelle 1: Definitionen von Nebeneffekten nach Linden (2013) ... 5
Tabelle 2: Häufigkeit von Nonresponse ... 6
Tabelle 3: Häufigkeit von Verschlechterung ... 6
Tabelle 4: Persönlichkeitsstörungscluster mit ihren zugehörigen Diagnosen und ICD-10-Codes nach DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5. Auflage) der American Psychiatric Association ... 13
Tabelle 5: Unterschiede zwischen der direkten und indirekten Veränderungserfassung nach Stieglitz und Baumann (2001) ... 19
Tabelle 6: Übersicht über die verwendeten Erhebungsinstrumente ... 26
Tabelle 7: Soziodemographische Merkmale mit ihren Ausprägungen ... 39
Tabelle 8: Soziodemographische Charakteristika der Untersuchungsstichprobe ... 44
Tabelle 9: ICD-10-Diagnosen der Patienten der Untersuchungsstichprobe ... 46
Tabelle 10: Vorbehandlung der Patienten ... 47
Tabelle 11: Therapeutenangaben zu Suizidalität, Suizidversuchen und parasuizidalem Verhalten der Patienten ... 47
Tabelle 12: Vergleich soziodemographischer Parameter der Teilstichprobe mit den Parametern der Gesamtstichprobe ... 49
Tabelle 13: Vergleich der Teil- und der Gesamtstichprobe anhand der zehn häufigsten Diagnosen der Teilstichprobe ... 51
Tabelle 14: Behandlungsergebnisse im Rahmen der indirekten Veränderungsmessung ... 54
Tabelle 15: Häufigkeit von Nonresponse und Verschlechterung in Abhängigkeit von der Messmethode ... 54
Tabelle 16: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach dem Geschlecht ... 55
Tabelle 17: Anzahl und Anteil der Patienten mit Erfolg und Misserfolg (inkl. Nonresponse) differenziert nach dem Geschlecht sowie Ergebnisse des χ^2-Tests ... 57
Tabelle 18: Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Geschlecht sowie Ergebnisse des χ^2-Tests ... 57
Tabelle 19: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Geschlechts ... 58
Tabelle 20: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach dem Vorliegen von Persönlichkeitsstörungen ... 61
Tabelle 21: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Erfolge und Misserfolge (inkl. Nonresponse) differenziert nach dem Vorliegen einer Persönlichkeitsstörung ... 63
Tabelle 22: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Vorliegen einer Persönlichkeitsstörung ... 63
Tabelle 23: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Vorliegens einer Persönlichkeitsstörung ... 64
Tabelle 24: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach den Persönlichkeitsstörungsclustern B und C ... 67

Tabelle 25: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Erfolge und Misserfolge (inkl. Nonresponse) differenziert nach dem Vorliegen einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung 69
Tabelle 26: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Vorliegen einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung 69
Tabelle 27: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Vorliegens einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung 70
Tabelle 28: Soziodemographische Aspekte differenziert nach dem Geschlecht 73
Tabelle 29: Soziodemographische Aspekte differenziert nach dem Vorliegen einer Persönlichkeitsstörung 75
Tabelle 30: Modell der logistischen Regression bezüglich des Einflusses soziodemographischer Aspekte auf das Auftreten von Therapiemisserfolg 77
Tabelle 31: Stichprobengröße, Frauenanteil und Altersdurchschnitt anderer Studien zu Misserfolgen in der stationären Psychotherapie 88
Tabelle 32: Behandlungsdauer und Therapiekonzept anderer Psychotherapiestudien 90
Tabelle 33: Relative Häufigkeiten von Patienten mit Persönlichkeitsstörungen und Diagnosen im Bereich der affektiven Störungen in anderen Studien 91
Tabelle 34: Ergebnisse der direkten Veränderungsmessung in anderen Studien 95
Tabelle 35: Ergebnisse der indirekten Veränderungsmessung in anderen Studien 98
Tabelle 36: Mittelwerte und Effektstärken der Summenscores der SCL-90-R, des BSS und des IIP in anderen Studien 102

Abkürzungsverzeichnis

AGGR	Skala Aggressivität/Feindseligkeit
ANGS	Skala Ängstlichkeit
BPtK	Bundespsychotherapeutenkammer
BSI	*Brief Symptom Inventory*
BSS	Beeinträchtigungs-Schwere-Score
c	*Cut-off*-Wert
d	Effektstärke
DEPR	Skala Depressivität
Destatis	Statistisches Bundesamt
df	Freiheitsgrade
DGKJP	Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie und -psychotherapie
DGP	Deutsche Gesellschaft für Psychologie
DGPM	Deutsche Gesellschaft für Psychosomatische Medizin und Ärztliche Psychotherapie
DGPPN	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde
DKPM	Deutsches Kollegium für Psychosomatische Medizin
DSM-5	*Diagnostic and Statistical Manual of Mental Disorders*, 5. Auflage
DSM-III	*Diagnostic and Statistical Manual of Mental Disorders*, 3. Auflage
e^b	Effektkoeffizient
GSI	*Global Severity Index*
ICC	*intraclass correlation coefficient*
ICD-10-GM	*International Statistical Classification of Diseases And Related Health Problems, 10th revision, German Modification*
IIP	Inventar zur Erfassung interpersonaler Probleme
IIP_{ges}	64-Item-Gesamtwert
kö	körperlich
PARA	Skala Paranoides Denken

PHOB	Skala Phobische Angst
PSKB-Se-R	Psychischer und Sozialkommunikativer Befund (Selbst/Revidiert)
psy	psychisch
PSYC	Skala Psychotizismus
QS	Quadratsumme
RC	*Reliable Change Index*
r_{tt}	Retest-Reliabilität
SCL	Symptom-Checkliste
SCL-90-R	Symptom-Checkliste (R = *revised*)
SCL-90-S	*Symptom-Checklist (S = Standard)*
soko	sozialkommunikativ
SOMA	Skala Somatisierung
S_v	Stanine
t	t-Wert
TER	*treatment emergent reactions*
UNSI	Skala Unsicherheit im Sozialkontakt
x_{jk}	unabhängige Variable
z_k	aggregierte Einflussstärke
ZWAN	Skala Zwanghaftigkeit
β	Regressionskoeffizient
$\eta_{partial}^2$	Effektstärke

1 Einleitung

Die stationäre Psychotherapie ist in Deutschland eine eigenständige, häufig angewandte und anerkannte Therapieform bei Erwachsenen, deren Wirksamkeit belegt ist (Liebherz und Rabung 2013; Spitzer et al. 2016; Wissenschaftlicher Beirat Psychotherapie 2005). Die Forschungsbestrebungen zu Misserfolgen in der stationären Psychotherapie in Deutschland hingegen sind geringer (Spitzer et al. 2012) und werden durch methodische Probleme, z. B. nicht ausreichende theoretische Konzeptualisierungen von Misserfolg, uneinheitliche Begriffsdefinitionen (Spitzer et al. 2012) oder fehlende Standards bei der Veränderungsmessung (Stieglitz und Hiller 2014), erschwert. Dementsprechend fehlen multimodal erhobene, differenzierte Angaben bezüglich der Häufigkeit von Nonresponse und Verschlechterungen sowie hinsichtlich möglicher Faktoren, die das Auftreten von Nonresponse und Verschlechterungen in der stationären psychotherapeutischen Behandlung bedingen. Aufgrund der in Deutschland speziellen Struktur der stationären psychotherapeutischen Versorgung können Ergebnisse aus dem ambulanten Sektor oder aus dem Ausland jedoch nicht problemlos auf die stationäre Behandlungssituation in Deutschland übertragen werden (Reuter et al. 2014).

Neben dem individuellen Leid für die Betroffenen, das sich aus einer wirkungslosen oder gar mit Verschlechterungen einhergehenden stationären Psychotherapie ergibt, sind auch gesundheitsökonomische Aspekte relevant. Die Krankheitskosten für Patienten[1] mit Diagnosen, die dem Kapitel V „Psychische und Verhaltensstörungen“ (F00-F99) der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10-GM) zuzuordnen sind, beliefen sich 2015 in Deutschland für die stationäre und teilstationäre Behandlung auf rund 24,9 Mrd. Euro (Statistisches Bundesamt (Destatis) 2017b). Das Erkennen oder Vorhersagen von erfolglosen Therapieverläufen in der psychotherapeutischen Behandlung ist somit auch aus ökonomischen Gründen sinnvoll. Darüber hinaus könnten einzelne der schätzungsweise 450 000 Patienten, die jährlich in Deutschland stationär psychotherapeutisch behandelt werden (Spitzer et al. 2016), davon profitieren, wenn eine Auswertung des Therapieverlaufs auf Einzelfallebene und das Wissen um Faktoren, die häufiger mit Misserfolgen einhergehen, dazu beitragen, dass eine Therapie optimiert werden kann. Wie häufig Misserfolge in der stationären Psychotherapie vorkommen und ob sich die Häufigkeit von Nonresponse und Verschlechterungen bei Männern und Frauen oder bei Patienten mit und ohne Persönlichkeitsstörungen oder nach der Art der Persönlichkeitsstörung unterscheiden, ist Thema dieser Arbeit.

[1] Auf Grund der besseren Lesbarkeit wird im Folgenden auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Die ausschließliche Verwendung der männlichen Form ist als geschlechtsunabhängig aufzufassen, sodass sämtliche Personenbezeichnungen für alle Geschlechter gelten.

In diesem Kapitel werden die Charakteristika der stationären Psychotherapie vorgestellt, der Forschungsstand hinsichtlich der Misserfolge in der Psychotherapie aufgezeigt und Grundlagen der Veränderungsmessung, der multimodalen Diagnostik und der Persönlichkeitsstörungen dargelegt.

1.1 Stationäre Psychotherapie

Im Folgenden werden die Versorgungsstruktur, die Häufigkeit, spezifische Charakteristika, Indikationen, Kontraindikationen und Therapiekonzepte der stationären Psychotherapie vorgestellt. Charakteristisch für die Versorgungssituation in Deutschland ist, dass neben der ambulanten Psychotherapie die stationäre Therapie psychischer und psychosomatischer Erkrankungen in einem der folgenden Versorgungsbereiche durchgeführt werden kann: in Krankenhäusern mit Fachabteilungen für Psychiatrie und Psychotherapie bzw. Fachabteilungen für Psychosomatik und Psychotherapie, die der früher verwendeten Gebietsbezeichnung „Psychotherapeutische Medizin“ entsprechen (Schulz et al. 2008), in Fachkliniken oder in psychosomatischen Rehabilitationseinrichtungen (Barghaan et al. 2007; Schulz et al. 2011).

Die Betten- und Fallzahlen der Fachabteilungen für Psychiatrie und Psychotherapie und für Psychosomatik und Psychotherapie der Krankenhäuser sind im Zeitraum von 2004 bis 2016 deutlich gestiegen. 2016 wurden deutschlandweit insgesamt 823 182 Patienten (2004: 712 533 Patienten) bei 55 976 bestehenden Betten (2004: 53 021 Betten) in Fachabteilungen für Psychiatrie und Psychotherapie und 84 600 Patienten (2004: 35 310 Patienten) bei 10 857 Betten (2004: 4 412 Betten) in Fachabteilungen für Psychosomatik und Psychotherapie in Krankenhäusern behandelt (BPtK 2014; Schulz et al. 2008; Statistisches Bundesamt (Destatis) 2017a). Die Anzahl der Patienten, die stationär eine Psychotherapie erhalten, wird dabei auf 450 000 innerhalb eines Jahres bei ca. 40 000 psychotherapeutischen Behandlungsplätzen in Krankenhäusern und Rehabilitationseinrichtungen geschätzt (Spitzer et al. 2016), wovon auf letztgenannte 2004 und 2007 circa 125 000 Fälle entfielen (Barghaan et al. 2009; Schulz et al. 2008).

In Abgrenzung zur ambulanten Psychotherapie gilt die stationäre Behandlung als eigenständige Therapieform (Hartkamp und Hildenbrand 2009; Spitzer et al. 2016). Diese zeichnet sich durch Multipersonalität, Multimodalität, Multiprofessionalität und eine aktive Milieugestaltung aus. Diese Charakteristika werden im Folgenden erklärt.

Die Multipersonalität beschreibt, dass der einzelne Patient während seiner stationären Therapie mit einer Gruppe von Mitpatienten und einer Gruppe von Therapeuten in Kontakt tritt (Spitzer et al. 2016).

Die Multimodalität ist dadurch gekennzeichnet, dass unterschiedliche Therapieangebote, die vor allem nichtsprachliche Zugangswege nutzen, ergänzend eingesetzt werden. Dazu zählen beispielsweise die Kunst-, Gestaltungs-, Ergo- und Musiktherapie (Spitzer et al.

2016). Außerdem werden „Paar- und Familiengespräche, Entspannungsverfahren, Biofeedback, körperliche Aktivierung, Training sozialer und emotionaler Kompetenzen oder metakognitiver Fähigkeiten, Selbstbeobachtungsverfahren, Psychoedukation, traumatherapeutische Methoden sowie Krankenpflege, v. a. Bezugspflege, aber natürlich auch leitlinienorientierte Psychopharmakotherapie“ (Spitzer et al. 2016: 101) im Rahmen eines multimodalen Therapiekonzeptes angewandt.

Unter Multiprofessionalität zur Methodenintegration in einen Gesamtbehandlungsplan wird verstanden, dass das multiprofessionelle Behandlungsteam Therapieziele und -schwerpunkte erarbeitet und evaluiert, nach denen sich der Einsatz, die Zusammenstellung, die Reihenfolge, die Intensität und die inhaltliche Gestaltung der Therapieangebote richten. Diese Aspekte sollen für jeden Patienten in einem Gesamtbehandlungsplan festgehalten werden (Spitzer et al. 2016).

Die aktive Milieugestaltung umfasst, dass das Krankenhaus und die Station als „soziale[r] Mikrokosmos“ (Spitzer et al. 2016: 101) und Wirkfaktor aktiv gestaltet werden sollten, sodass sie therapeutisch erwünschte Veränderungen ermöglichen und fördern (Spitzer et al. 2016).

Dadurch, dass sich die stationäre Psychotherapie hinsichtlich der oben dargestellten Aspekte deutlich von der ambulanten Versorgung unterscheidet, müssen auch die Indikationen und Kontraindikationen spezifisch für die stationäre Behandlung formuliert werden. Eine stationäre psychotherapeutische Behandlung gilt somit als indiziert, wenn „das Behandlungsziel nicht durch teilstationäre, vor- und nachstationäre oder ambulante Behandlung einschließlich häuslicher Krankenpflege erreicht werden kann“ (Sozialgesetzbuch, Fünftes Buch, Gesetzliche Krankenversicherung 2019: § 39 Absatz 1 Satz 2) und kommt u. a. zur Anwendung bei Depressionen, den Essstörungen Bulimie und Anorexie, bei dissoziativen und somatoformen Störungen, Anpassungs-, Angst-, Zwangs- und Persönlichkeitsstörungen und posttraumatischen Belastungsstörungen (Beutel und Subic-Wrana 2010; Spitzer et al. 2016). Dabei ist zur Indikationsstellung auch die Präferenz des Patienten nach umfassender Aufklärung über sein Krankheitsbild und dessen Behandlungsmöglichkeiten zu berücksichtigen (Spitzer et al. 2016). Hirnorganische Erkrankungen, akute Psychosen, akute Selbst- und Fremdgefährdung und Suchterkrankungen gelten auf offenen Stationen meist als Kontraindikation (Beutel und Subic-Wrana 2010).

In der stationären Psychotherapie werden unterschiedliche Therapiekonzepte angewandt. Es können psychodynamische von verhaltenstherapeutischen und integrativen Ansätzen abgegrenzt werden (Spitzer et al. 2016). Während auf die verhaltenstherapeutischen Ansätze hier nicht näher eingegangen wird, da sie für diese Arbeit kaum relevant sind, sollen zunächst der Begriff und die Grundannahmen der psychodynamischen Psychotherapie erläutert werden. Zur psychodynamischen Psychotherapie gehören die analytischen und die tiefenpsychologisch fundierten Psychotherapien (Wissenschaftlicher Beirat Psychotherapie 2005), die vor der Einführung des Oberbegriffs der psychodynamischen Psychotherapie als

„psychoanalytisch begründete“ Verfahren bezeichnet wurden (Rudolf und Rüger 2012). Sie stellen modifizierte Anwendungen der klassischen Psychoanalyse dar (Beutel et al. 2010; Rüger und Reimer 2012; Wissenschaftlicher Beirat Psychotherapie 2005). Die einzelnen Verfahren der psychodynamischen Psychotherapie wie die analytische, die tiefenpsychologisch fundierte, die dynamische oder die psychoanalytisch-interaktionelle Psychotherapie (Reimer und Rüger 2012) unterscheiden sich hinsichtlich der Strukturierung, der Interventionstechnik des Therapeuten oder der Fokussierung auf aktuelle oder frühere Konflikte (Wissenschaftlicher Beirat Psychotherapie 2005). Die Ausprägungen dieser Eigenschaften, in denen sie sich teilweise unterscheiden, liegen dabei auf einem Kontinuum, sodass die klare Abgrenzbarkeit der einzelnen Verfahren voneinander eingeschränkt ist und sich die Übergänge zwischen den Verfahren fließend gestalten (Rudolf und Rüger 2001). In der psychodynamischen Psychotherapie wird angenommen, dass eine für einen Patienten typische dysfunktionale Beziehungsgestaltung seine inneren Konflikte, seine strukturellen Einschränkungen und seine gewohnten Objektbeziehungen widerspiegelt (Spitzer et al. 2016). Unter der Beachtung von Übertragung und Gegenübertragung (Spitzer et al. 2016) werden dementsprechend in der psychodynamischen Psychotherapie psychosoziale Konflikte bearbeitet (Rüger und Reimer 2012). Hierbei ist das Dreieck aus der interpersonellen Beziehung zwischen Therapeut und Patient, der symptomauslösenden oder -verstärkenden interpersonellen Situation und den pathogenen Interaktionsmustern im sozialen Umfeld Gegenstand der Therapie (Heigl-Evers et al. 1997). Unter Zuhilfenahme dieses „modifizierten Dreiecks der Einsichtsvermittlung“ (Heigl-Evers et al. 1997: 158) soll der Patient Einsicht in seine inneren Konflikte bekommen, sich innerlich umstrukturieren und eine rückläufige Symptomatik aufweisen (Heigl-Evers et al. 1997).

Neben den streng abgegrenzten verhaltenstherapeutischen und psychodynamischen Therapiekonzepten existieren integrative Ansätze. Der Begriff der Integration bezieht sich bei der stationären integrativen Psychotherapie darauf, dass sie schulen- und methodenübergreifend arbeitet, die Trennung von Real- und Therapieraum aufhebt, ggf. eine medikamentöse Therapie einschließt und Alltagsaufgaben wie die Essensversorgung in den Gesamtbehandlungsplan eingliedert (Spitzer et al. 2016).

1.2 Misserfolge in der Psychotherapie

In diesem Kapitel werden die für diese Arbeit relevanten Begriffe „Nonresponse“ und „Verschlechterung“ im Kontext der Definitionen von Nebeneffekten, Nebenwirkungen und Misserfolgen eingeführt. Im Anschluss daran wird der Forschungsstand bezüglich der Fragen, wie häufig Nonresponse und Verschlechterungen in der stationären Psychotherapie vorkommen und welche Variablen als Prädiktoren für Nonresponse und Verschlechterungen gelten, dargestellt.

1.2.1 Abgrenzungen der Begrifflichkeiten in der Misserfolgsforschung

Einige der Begriffe, die für die Misserfolgsforschung relevant sind, werden nicht klar voneinander differenziert verwendet, da eine einheitliche Definition bisher nicht existiert, obwohl bereits in den 1970er und 1980er Jahren u.a. eine Arbeitsgruppe um Strupp (1977; 1980) negative Effekte von Psychotherapie untersuchte. Einen aktuellen Systematisierungsversuch liefert Linden (2013), der die in Tabelle 1 dargestellten Begrifflichkeitsabgrenzungen und Definitionen vorschlägt.

Tabelle 1: Definitionen von Nebeneffekten nach Linden (2013)

Begriff deutsch	Begriff englisch	Definition
unerwünschtes Ereignis	Unwanted event (UE)	jegliche negative Ereignisse zeitgleich zur Therapie
negative Therapiefolgen	Treatment emergent reactions (TER)	durch die Behandlung verursachte UE
Nebenwirkungen	Adverse treatment reactions (ATR)	TER bei korrekt angewandter Therapie
Kunstfehlerfolgen	Malpractice reaction (MPR)	TER bei fehlerhaft angewandter Therapie
Therapie-Nonresponse	Treatment nonresponse (TNR)	ungenügende Besserung trotz Therapie
Krankheitsverschlechterung	Deterioration of illness (DOI)	Verschlechterung einer Erkrankung trotz Therapie
therapeutische Risiken	Therapeutic risk (TR)	bekannte und vorhersehbare Nebenwirkungen
Kontraindikationen	Contraindications	Bedingungen, die das Auftreten von Nebenwirkungen wahrscheinlich machen

Der Begriff der Therapieschäden kommt durch die genauere Differenzierung des Begriffs der negativen Therapiefolgen (s. Tabelle 1) zustande und kann diese Liste an Begriffen ergänzen. Unter Therapieschäden werden unvorhergesehene, andauernde, für den Patienten negative Therapiefolgen gefasst (Hoffmann et al. 2008).

Neben diesem Systematisierungsversuch existieren mehrere Annäherungen an den Begriff der Nebenwirkungen. Nebenwirkungen von Psychotherapie sind multidimensional (Lilienfeld 2007) und können sich in Bereichen manifestieren, die über das bestehende Beschwerdebild hinausgehen, z. B. als Veränderungen der Arbeitsfähigkeit oder der sozialen Teilhabe oder im Auftreten von Beziehungsproblemen (Haupt et al. 2013). Jacobi (2001) definiert jedoch Nebenwirkungen eng als „vorhersehbare, klar umgrenzte und kurzfristige Stimmungsveränderungen“ (Jacobi 2001: 328) im Rahmen der Therapie. Haupt et al. (2013) stellen bezüglich des Begriffs der Nebenwirkungen zwei Ansätze gegenüber. Während der eine Ansatz Nebenwirkungen als genauer differenzierbare Fehlanwendungen einer Therapie auffasst, setzt der andere Ansatz Nebenwirkungen mit Misserfolgen gleich (Haupt et al.

2013). Misserfolge werden als „die unzureichende oder ganz fehlende Erreichung der in der Therapie explizit und implizit angestrebten Ziele“ (Hoffmann et al. 2008: 5) verstanden. Bei Misserfolgen kann laut Jacobi (2001) weiter zwischen folgenden Aspekten differenziert werden: Misserfolge vor Therapiebeginn, Patienten, die eine Therapie ablehnen, Therapieabbrüche, Nonresponse, Rückfälle oder langfristige Verschlechterungen. Im weiteren Verlauf der Arbeit stehen Nonresponse und Verschlechterungen im Vordergrund.

1.2.2 Häufigkeit von Nonresponse und Verschlechterung

Den Stand der Misserfolgsforschung in der stationären psychotherapeutischen Krankenhausbehandlung – basierend auf einer Literaturrecherche in den Datenbanken MEDLINE, PsycARTICLES, PsycINFO, und PSYNDEX – stellen Reuter und Kollegen (2014) in einem Übersichtsartikel dar. Die Angaben der Häufigkeit von Nonresponse und Verschlechterung zum Zeitpunkt der Entlassung variieren je nach Studie und Art der Verlaufsmessung (s. Tabelle 2 und Tabelle 3). Die meisten Studien berichten für die Häufigkeit von Nonresponse einen Anteil von 20 bis 30 % und für Verschlechterungen einen Anteil von 3 bis 10 % der Patienten (Reuter et al. 2014).

Tabelle 2: Häufigkeit von Nonresponse

		Nonresponse (%)	Größe der Stichprobe und Quelle
Direkte Verlaufsmessung	Minimalwert	4,9	n = 164 (Junge und Ahrens 1996)
	Maximalwert	24,2	n = 741 (Kriebel et al. 2001)
Indirekte Verlaufsmessung	Minimalwert	4,9	n = 164 (Junge und Ahrens 1996)
	Maximalwert	44,3	n = 43 (Geiser et al. 2001)
	meist gemessene Werte	20-30	(Reuter et al. 2014)

Tabelle 3: Häufigkeit von Verschlechterung

		Verschlechterung (%)	Größe der Stichprobe und Quelle
Direkte Verlaufsmessung	Minimalwert	0,5	n = 1310 (Rief et al. 1994)
	Maximalwert	13,8	n = 741 (Kriebel et al. 2001)
Indirekte Verlaufsmessung	Minimalwert	1,8	n = 164 (Junge und Ahrens 1996)
	Maximalwert	18,8	n = 213 (Sammet et al. 2004)
	meist gemessene Werte	3-10	(Reuter et al. 2014)

1.2.3 Korrelate von Nonresponse und Verschlechterung

In der Literatur werden soziodemographische, sozialmedizinische, klinische und therapiebezogene Korrelate bzw. Prädiktoren des Therapieerfolges diskutiert.

Zu den untersuchten soziodemographischen Parametern gehören das Geschlecht, das Bestehen einer Partnerschaft und das Vorliegen bzw. die Höhe eines Schulabschlusses. Die Frage, ob das Geschlecht einen Einfluss auf die Häufigkeit von Therapiemisserfolgen hat, wird widersprüchlich beantwortet (Bassler 1995; Geiser et al. 2001; Geyer et al. 2003; Konzag et al. 2004; Kriebel et al. 2001; Spitzer et al. 2012). Keinen Einfluss des Geschlechts auf die Häufigkeit von Nonresponse oder Verschlechterungen fanden Geyer und Kollegen (2003) in ihrer Diskriminanzanalyse (Stichprobengröße n = 61), Kriebel und Kollegen (2001) (n = 741), Bassler (1995) (n = 218) sowie die Gruppe um Konzag (2004) (n = 254 Patienten mit Persönlichkeitsstörungen). Reuter und Mitarbeiter (2014) zählten zwei Studien auf, in denen Männer ein niedrigeres Risiko für Nonresponse und Verschlechterung hätten. Für eine der beiden Studien (n = 495) trifft diese Aussage nur für die Verschlechterung, nicht jedoch für Nonresponse zu: 5 % der Männer und 6 % der Frauen verschlechterten sich im Rahmen der Therapie, jedoch waren 24 % der Männer und 21 % der Frauen in dieser Multi-Center-Studie Nonresponder (Franz et al. 2000). In der zweiten Studie verschlechterte sich keiner der 12 Männer unter den 43 untersuchten Patienten (Geiser et al. 2001).

Die Arbeitsgruppe um Geiser (2001) beschrieb für das Bestehen einer Partnerschaft, dass in ihrer Studie (n = 43) unter den geheilten Patienten, also unter denen, die sich statistisch und klinisch signifikant verbesserten, 56 % verheiratet waren und unter den gebesserten, die sich nur statistisch signifikant verbesserten, 11 %. Für die Gruppe der Nonresponder und der Verschlechterten trafen sie keine Aussage. Auch die Gruppe um Geyer (2003) (n = 61) stellte heraus, dass Patienten, die in Partnerschaften oder Familien leben, häufiger von Psychotherapie profitieren. Bassler (1995) und Kriebel (2001) (n = 741) konnten bezüglich des Familienstandes hingegen keinen signifikanten Zusammenhang feststellen.

In Hinblick auf den Schulabschluss stellen sich die Ergebnisse kontrovers dar: In der Studie von Bassler (1995) korreliert die Schulbildung negativ mit dem Therapieerfolg (n = 218). Die Patienten, die erfolgreich therapiert wurden, wiesen häufiger keinen Schulabschluss, einen Sonderschul- oder einen Hauptschulabschluss auf als die Gruppe der nicht erfolgreich Therapierten. Während Patienten mit Realschulabschluss in dieser Studie seltener erfolgreich therapiert wurden, konnte das Abitur nicht als prognostischer Faktor herangezogen werden (Bassler 1995). Die Therapiemotivation aus Therapeutensicht nahm jedoch mit steigendem Bildungsgrad signifikant zu (Bassler 1995). Bassler (1995) widersprechend befinden sich Patienten mit höher qualifizierten Berufen laut Geyer und Kollegen (2003) häufiger in der Gruppe der erfolgreich Therapierten. Kriebel und Mitarbeiter (2001) hingegen fanden keinen Zusammenhang zwischen dem Schulabschluss und dem Therapieergebnis (n = 741, davon verschlechtert: n = 102, unverändert: n = 179, verbessert: n = 460).

Sozialmedizinische Faktoren, die untersucht wurden, sind die Art der Erwerbstätigkeit, das Vorliegen einer Arbeitsunfähigkeit vor der stationären Aufnahme und das Bestehen einer

Rente oder eines Rentenantrags. Die Erwerbstätigkeit hat den Studien von Bassler (1995) und Kriebel und Mitarbeitern (2001) zufolge keinen signifikanten Effekt auf den Therapieerfolg.

Laut Kriebel und Kollegen (2001) ist in der Gruppe der Verschlechterten der Anteil der Patienten, die zu Therapiebeginn arbeitsunfähig waren, höher. Bezüglich der Länge der Arbeitsunfähigkeit differenzierten sie weiter: Patienten, die vor Therapiebeginn maximal drei Monate krankgeschrieben waren, gehörten empirisch häufiger zur Gruppe der verbesserten Patienten, als bei statistischer Unabhängigkeit zu erwarten wäre. Patienten, die über sechs Monate krankgeschrieben waren, gehörten hingegen seltener als statistisch erwartet zu der Gruppe der verschlechterten Patienten (Kriebel et al. 2001). Die Regressionsanalysen von der Arbeitsgruppe um Oster (2009) (n = 463) ergaben, dass die Arbeitsunfähigkeitsdauer in dem Jahr vor einer psychosomatischen Rehabilitation signifikant negativ mit einem der folgenden drei Aspekte korreliert: mit einer Besserung des Befindens drei Monate nach Entlassung oder mit der Verbesserung sozialmedizinischer Faktoren zum Entlassungszeitpunkt oder mit einer Entlassung im arbeitsfähigen Zustand.

Eine Berentung oder ein laufendes Rentenverfahren beeinflusst nach Kriebel und Kollegen (2001) die Häufigkeit der Misserfolge nicht signifikant. Der Arbeitsgruppe um Oster (2009) zufolge korreliert ein Rentenbegehren oder eine Rentenempfehlung durch einen Arzt jedoch signifikant negativ mit den Erfolgskriterien.

Beim Einfluss klinischer Faktoren auf das Auftreten von Nonresponse und Verschlechterungen wurden die Diagnosen, die Beeinträchtigungsschwere, das Vorhandensein interpersoneller Probleme, die Dauer der Erkrankung, psychotherapeutische Vorbehandlungen und Suizidalität als Variablen untersucht. Zur Korrelation zwischen den Diagnosen und Nonresponse oder Verschlechterungen liegen die im Folgenden aufgezählten Studien vor. Rudolf und seine Mitarbeiter (2004) stellten heraus, dass sich der größte Anteil nicht gebesserter Patienten bei jenen mit somatoformen Störungen finden ließ (37 % von 65 Patienten). Weniger Nonresponse oder Verschlechterungen gaben sie für Patienten mit affektiven Störungen (23 % von 172 Patienten), mit histrionischen, ängstlich abhängigen oder anankastischen Persönlichkeitsstörungen (22 % von 102 Patienten), mit Essstörungen (21 % der 121 Patienten) oder mit schizoider, emotional instabiler oder narzisstischer Persönlichkeitsstörung (19 % von 107 Patienten) an. Die Rate der nicht Gebesserten war in der Gruppe der Patienten mit Angst- und Zwangsstörungen mit 15 % (von 107 Patienten) am niedrigsten (Rudolf et al. 2004). Bei den Persönlichkeitsstörungen differenzierten Spitzer und seine Mitarbeiter (2012) nach Clustern (s. 1.2.4). Die höchsten Misserfolgsraten fanden sie bei Patienten mit Cluster-B- oder kombinierten Persönlichkeitsstörungen. Je nach Art der Verlaufsmessung gaben sie für die Cluster-B-Persönlichkeitsstörungen (n = 555) Misserfolgsraten von 12,8 %, ermittelt durch eine direkte Veränderungsmessung aus Therapeutensicht, bis 32,0 %, gemessen mithilfe der indirekten Veränderungsmessung aus Patientenperspektive anhand der Symptomcheckliste SCL-90-R (s. 2.3.1), an. Mit denselben Veränderungs-

messungen fanden sie für kombinierte Persönlichkeitsstörungen (n = 135) Misserfolgsraten von 16,7 % bis 34,2 % (Spitzer et al. 2012). Franz und Kollegen (2000) gaben in einer Multi-Center-Studie ca. 40-50 % Nonresponse und Verschlechterung bei Patienten mit Borderline-Störung (n = 4), für Patienten mit Panikstörung (n = 7) und für Patienten mit Bulimie (n = 7) an.

Eine weitere klinische Variable, deren Einfluss auf Nonresponse und Verschlechterungen erforscht wurde, ist die Beeinträchtigungsschwere. Eine hohe Beeinträchtigungsschwere aus Therapeutenperspektive (Bassler 1995; Geyer et al. 2003; Oster et al. 2009) und aus Patientensicht (Kriebel et al. 2001; Schauenburg et al. 2001) korreliert mit Misserfolg oder Nonresponse in der Therapie. Die Arbeitsgruppe um Konzag (2004) fand jedoch keinen Zusammenhang zwischen der Beeinträchtigungsschwere aus Therapeutensicht und dem Therapieerfolg aus der Patientenperspektive.

Darüber hinaus wurden interpersonelle Probleme als klinische Variable analysiert. Patienten, deren Krankheitsbild durch eine Therapie verbessert oder geheilt wurde, berichteten zu Therapiebeginn häufiger von schwereren interpersonellen Problemen als Patienten in der Gruppe der Verschlechterten. Vor allem in den Bereichen Unterwürfigkeit, Konkurrenz und Introversion, welche durch das Inventar zur Erfassung interpersonaler Probleme (IIP) (s. 2.3.2) betrachtet werden, wiesen die geheilten und verbesserten Patienten signifikant höhere Beeinträchtigungen vor der Therapie auf (Davies-Osterkamp et al. 1996). Im Verlauf der Therapie nahmen die interpersonellen Probleme in der Gruppe der Patienten, deren generelle psychische Belastung sich signifikant gebessert hatte, signifikant ab (n = 129) (Sammet et al. 2004). In der Gruppe der Patienten mit unveränderter allgemeiner Symptombelastung (n = 44) veränderten sich die interpersonellen Probleme ebenfalls nicht signifikant und in der Gruppe der signifikant verschlechterten Patienten (n = 40) nahmen sie signifikant zu (Sammet et al. 2004). Während Sammet und Kollegen (2004) wie beschrieben Patientengruppen nach den Ergebnissen der SCL-90-R einteilten und deren interpersonelle Probleme untersuchten, differenzierten Haase und Mitarbeiter (2008) vier Gruppen bezüglich der Veränderung des IIP im Verlauf und untersuchten deren Symptombelastung. Die vier Gruppen bezüglich der Veränderung des IIP sind die Gruppe der klinisch und statistisch verbesserten Patienten (n = 70), eine Gruppe mit bei Aufnahme unauffälligen IIP-Werten (n = 57), eine Gruppe der Unveränderten (n=19) und eine Gruppe mit Patienten, die sich im IIP verschlechterten (n = 10). Die im IIP verschlechterten oder unveränderten Patienten wiesen im Verlauf der Therapie und zum Nacherhebungszeitpunkt nach einem Jahr keine signifikante Verbesserung der Symptombelastung in der SCL-90-R auf (Haase et al. 2008).

Der klinischen Variable „Dauer der Erkrankung“ wird zum Teil keine Korrelation mit Nonresponse zugeschrieben (Geiser et al. 2001; Kriebel et al. 2001; Schauenburg et al. 2001), z. T. jedoch auch ein positiver Zusammenhang (Geyer et al. 2003). Bassler (1995) fand diesbezüglich eine schwache Korrelation (n = 218) und inferenzstatistisch einen

„knapp tendenzielle[n] Unterschied“ (Bassler 1995: 83) (n = 212, Behandlungsdauer der erfolgreichen Gruppe Mittelwert M = 59,0 ± 76,8 Monate vs. M = 75,5 ± 91,9 Monate in der erfolglosen Gruppe).

Inwiefern die klinische Variable der psychotherapeutischen Vorbehandlung einen Einfluss auf Nonresponse und Verschlechterungen hat, wurde in der Literatur wie im Folgenden beschrieben diskutiert. Bassler (1995) bewertete Vorbehandlungen als „tendenziell bedeutsam“ (Bassler 1995: 84). Psychotherapeutisch vorbehandelte Patienten weisen demzufolge häufiger erfolglose Therapieverläufe auf als Patienten ohne Psychotherapie-Erfahrungen (Bassler 1995). Signifikante Unterschiede zwischen verbesserten, unveränderten und verschlechterten Patienten fand die Arbeitsgruppen um Kriebel (2001). Geyer und Kollegen (2003) zufolge hat eine psychotherapeutische Vorbehandlung eine niedrige prognostische Bedeutung.

Außerdem wurde Suizidalität als klinischer Faktor, der im Zusammenhang mit Nonresponse und Verschlechterungen steht, untersucht. Patienten ohne Behandlungserfolg waren vor Behandlungsbeginn häufiger suizidal als erfolgreich behandelte Patienten (61,6 % vs. 53,7 %) (Spitzer et al. 2012).

Studien zum Einfluss therapiebezogener Faktoren untersuchten u.a. die Therapiebeziehung und die Behandlungsdauer als Variablen. Eine schlechtere therapeutische Beziehung korreliert mit Misserfolg (Bassler 1995; Konzag et al. 2004; Liebler et al. 2004; Sammet et al. 2004). Bezüglich der Korrelation zwischen der Behandlungsdauer und der Misserfolgshäufigkeit konstatierten Schauenburg und Mitarbeiter (2001) in ihrer Studie (n = 71), dass die Behandlungsdauer an den Verlauf der Besserung adaptiert werde, wobei die Patienten, die sich schnell verbessern, früher entlassen werden. Dies deckt sich mit der Beobachtung von Geiser und Kollegen (2001) (n = 60), dass die mittlere Behandlungsdauer in der Gruppe der geheilten Patienten mit neun Wochen kürzer war als in den Gruppen der gebesserten, unveränderten oder verschlechterten Patienten mit circa elf Wochen (Geiser et al. 2001) und mit den Daten von Spitzer und Kollegen (2012) (n = 1239, 104,2 ± 27,7 Tage in der Misserfolgsgruppe vs. 100,9 ± 23,0 Tage in der Gruppe der erfolgreich Therapierten).

Zusammenfassend lässt sich festhalten, dass die Studienergebnisse für die Variablen, die in diesem Kapitel beschrieben werden, mit Ausnahme der Variablen Erwerbstätigkeit, Arbeitsunfähigkeit vor der stationären Aufnahme, Therapiebeziehung und Behandlungsdauer keine übereinstimmenden Ergebnisse hinsichtlich des Zusammenhangs mit Nonresponse und Verschlechterungen ergeben.

Die in Kapitel 1.2.2 und in diesem Kapitel genannten Studien verwendeten teilweise sehr kleine Stichproben, sodass ihre Repräsentativität und Aussagekraft stark eingeschränkt ist (z. B. n = 43 (Geiser et al. 2001), n = 61 (Geyer et al. 2003); weitere Stichprobengrößen s. Tabelle 31). Besonders problematisch ist die Betrachtung von Untergruppen solcher kleinen Stichproben. Beispielsweise kann die Angabe von 50 % Nonresponse und Verschlech-

terung bei Patienten mit Borderline-Störung in der Studie von Franz und Kollegen (2000) nicht verallgemeinert werden, da sie sich auf eine zu kleine Stichprobe (n = 4) stützt.

Neben einer kritischen Betrachtung der Stichprobengröße ist auch auf die Zusammensetzung der Stichprobe zu achten. Teilweise ergeben sich durch die in den o. g. Studien formulierten Einschlusskriterien diagnostisch homogene Stichproben. Beispielsweise bezogen sich die Arbeitsgruppen um Konzag (2004) und Spitzer (2012) nur auf Patienten mit Persönlichkeitsstörungen. Die Ergebnisse der Studien an diagnostisch homogenen Stichproben können nicht ungeprüft auf diagnostisch heterogene Gruppen übertragen werden. Bei den Studien mit diagnostisch heterogenen Stichproben (Bassler 1995; Davies-Osterkamp et al. 1996; Franz et al. 2000; Geiser et al. 2001; Geyer et al. 2003; Haase et al. 2008; Junge und Ahrens 1996; Kriebel et al. 2001; Liebler et al. 2004; Nübling et al. 2004; Davies-Osterkamp et al. 1996; Oster et al. 2009; Rief et al. 1994; Sammet et al. 2004; Schauenburg et al. 2001) ist die Vergleichbarkeit der Zusammensetzung der Diagnosen nur eingeschränkt möglich, da sie in unterschiedlicher Weise angegeben wurden (s. 4.3.1.2). Durch die begrenzte Vergleichbarkeit der Stichproben ist auch die Vergleichbarkeit der Studienergebnisse eingeschränkt.

Beim Vergleich der Studiensettings fällt auf, dass die Mehrzahl der aufgeführten Studien vollständig an Universitätskliniken durchgeführt wurden (Bassler 1995; Franz et al. 2000; Geiser et al. 2001; Konzag et al. 2004; Liebler et al. 2004; Rudolf et al. 2004; Sammet et al. 2004) oder zumindest teilweise auf Daten aus Universitätskliniken basierten (Davies-Osterkamp et al. 1996; Oster et al. 2009). In der Studie von Geyer und Kollegen (2003) fehlen jegliche Angaben zur Art des Krankenhauses, in dem die Daten erhoben wurden. Durch Unterschiede in den Strukturen, klinischen Abläufen und Zielsetzungen von Universitäts- und Versorgungskrankenhäusern, also in psychiatrischen, psychosomatischen und psychotherapeutischen Fachkrankenhäusern der Regelversorgung, können beispielsweise auch Unterschiede im Patientenkollektiv oder der Behandlungsdauer und dem Behandlungsergebnis vorliegen, sodass eine Übertragung der Forschungsergebnisse aus den Universitätskliniken auf die Situation in Versorgungskrankenhäusern nicht problemlos möglich ist. Lediglich die Studien der Arbeitsgruppen um Haase (2008), Kriebel (2001), Spitzer (2012) und Nübling (2004) wurden statt in Universitätskrankenhäusern in Versorgungskrankenhäusern durchgeführt. Letztgenannte Studien unterscheiden sich zusätzlich darin, ob sie in Akut- oder Rehabilitationskliniken erhoben wurden. Vor dem Hintergrund, dass sich die Therapieziele dieser Einrichtungen in der Form unterscheiden, dass eine Rehabilitationsmaßnahme der Abwendung einer drohenden Erwerbsunfähigkeit dienen soll und die Krankenhausbehandlung auf eine „kurative[ ] Therapie einer krankheitswertigen Störung" (Spitzer et al. 2016: 100) zielt, ergibt sich, dass eine Berücksichtigung der Einrichtungsformen bei der Betrachtung der Studienergebnisse notwendig ist. Von den o. g. Studien in außeruniversitären Einrichtungen basieren jene von Kriebel und Kollegen (2001) und Nübling und Mitarbeitern (2004) auf Erhebungen in Rehabilitationseinrichtungen. Unter Berücksichtigung des Settings (Universität versus Versorgungsklinik; Rehabilitationseinrich-

tung versus Akutkrankenhaus) liegen somit lediglich zwei Studien (Haase et al. 2008; Spitzer et al. 2012) zu Nonresponse- und Verschlechterungen bei stationärer Psychotherapie in außeruniversitären Fachkliniken vor. Berücksichtigt werden sollte, dass in der Studie von Spitzer und Kollegen (2012) nur Patienten mit Persönlichkeitsstörungen untersucht wurden, sodass lediglich eine Studie (Haase et al. 2008) zu Nonresponse- und Verschlechterungen bei stationärer Psychotherapie in außeruniversitären Fachkliniken an einer diagnostisch heterogenen Stichprobe besteht.

1.2.4 Persönlichkeitsstörungen

Bei der Untersuchung des Zusammenhangs zwischen klinischen Faktoren und dem Auftreten von Nonresponse und Verschlechterungen wurde unter anderem der Einfluss der Diagnosen untersucht (s. 1.2.3). Hierbei wurde von der Arbeitsgruppe um Spitzer (2012) speziell auch der Zusammenhang mit Persönlichkeitsstörungen genauer beleuchtet. Obwohl 2015 in deutschen Krankenhäusern 36.970 Patienten mit Persönlichkeitsstörungen als Hauptdiagnose (Statistisches Bundesamt (Destatis) 2016a) behandelt wurden und Persönlichkeitsstörungen somit in der stationären Psychotherapie ein relevantes Krankheitsbild darstellen, existiert allein die genannte Studie von Spitzer und Kollegen (2012) zu Misserfolgen in der stationären Psychotherapie von Patienten mit Persönlichkeitsstörungen. Diese untersuchte bereits eine große Stichprobe (n = 1239) unter Beachtung der unterschiedlichen Arten der Veränderungsmessung (s. 1.3) (Spitzer et al. 2012). Da jedoch als Kriterium für Misserfolg ausschließlich die allgemeine Symptombelastung herangezogen wurde (Spitzer et al. 2012), ist noch nicht bekannt, inwiefern Misserfolge beispielsweise hinsichtlich interpersonaler Probleme in der stationären Psychotherapie vorliegen. Es ergibt sich somit ein Forschungsbedarf hinsichtlich Nonresponse und Verschlechterungen bei Patienten mit Persönlichkeitsstörungen. Auf diesen wird in dieser Arbeit eingegangen, weshalb im Folgenden die Persönlichkeitsstörungen kurz vorgestellt werden.

Als Persönlichkeitsstörungen gelten zeitlich stabil fortbestehende, tiefgreifende, unflexible Kombinationen aus innerem Erleben und Verhalten einer Person, die den Ansprüchen des Umfelds widersprechen, sich in der Adoleszenz oder im frühen Erwachsenenalter entwickeln und die Person beeinträchtigen oder bei ihr einen Leidensdruck entstehen lassen (Falkai et al. 2015). Die Persönlichkeitsstörungen lassen sich – einem kategorialen Ansatz folgend – in drei Gruppen einteilen (s. Tabelle 4) (Falkai et al. 2015). In Cluster A (sonderbar-exzentrisch) werden die Paranoide, die Schizoide und die Schizotype Persönlichkeitsstörung zusammengefasst. Die Antisoziale, die Borderline-, die Histrionische und die Narzisstische Persönlichkeitsstörung bilden das Cluster B (dramatisch-emotional). Cluster C (ängstlich-furchtsam) enthält die Vermeidend-Selbstunsichere, die Dependente und die Zwanghafte Persönlichkeitsstörung. Diese Einteilung ist jedoch nicht umfassend validiert. Bei einer Person können auch Persönlichkeitsstörungen aus mehreren Clustern auftreten (Falkai et al. 2015).

Tabelle 4: Persönlichkeitsstörungscluster mit ihren zugehörigen Diagnosen und ICD-10-Codes nach DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5. Auflage) der American Psychiatric Association (Falkai et al. 2015)

ICD-10-Code	Persönlichkeitsstörung
Cluster A	
F60.0	Paranoide Persönlichkeitsstörung
F60.1	Schizoide Persönlichkeitsstörung
F21	Schizotype Persönlichkeitsstörung
Cluster B	
F60.2	Antisoziale Persönlichkeitsstörung
F60.3	Borderline Persönlichkeitsstörung
F60.4	Histrionische Persönlichkeitsstörung
F60.81	Narzisstische Persönlichkeitsstörung
Cluster C	
F60.6	Vermeidend-Selbstunsichere Persönlichkeitsstörung
F60.7	Dependente Persönlichkeitsstörung
F60.5	Zwanghafte Persönlichkeitsstörung
Andere Persönlichkeitsstörungen	
F07.0	Persönlichkeitsveränderung aufgrund eines anderen medizinischen Krankheitsfaktors
F60.89	Andere näher bezeichnete Persönlichkeitsstörung
F60.9	Nicht näher bezeichnete Persönlichkeitsstörung

Gleichzeitig besteht ein alternatives, dimensionales DSM-5-Modell für Persönlichkeitsstörungen. In diesem gelten Persönlichkeitsstörungen als „unangepasste Varianten von Persönlichkeitszügen mit fließenden Übergängen sowohl zur Normalität als auch zueinander“ (Falkai et al. 2015: 885). Persönlichkeitsstörungen werden in diesem Modell durch die Beurteilung der Einschränkungen im Funktionsniveau der Persönlichkeit, das die Selbst- und interpersonellen Funktionen umfasst, und durch die Einschätzung problematischer Persönlichkeitsmerkmale beschrieben. Die Selbstfunktion enthält die Aspekte Identität und Selbststeuerung. Empathie und Nähe bilden die interpersonelle Funktion. In dem alternativen DSM-5-Modell werden sechs spezifische Persönlichkeitsstörungen differenziert: die Antisoziale, Vermeidend-Selbstunsichere, Borderline-, Narzisstische, Zwanghafte und Schizotype Persönlichkeitsstörung. Diese werden von der Persönlichkeitsstörung, Merkmalsspezifiziert abgegrenzt (Falkai et al. 2015).

In der Allgemeinbevölkerung weist circa eine von sieben Personen eine Persönlichkeitsstörung auf (Eaton und Greene 2017). Bei einer Studie mit 452 Personen aus 109 Familien in Deutschland wurde bei 9,6 % der Männer und 10,3 % der Frauen im Laufe ihres Lebens mindestens eine Persönlichkeitsstörung nach DSM-III im Rahmen eines Interviews festgestellt (Maier et al. 1992). Unter psychiatrisch behandelten Patienten liegt der Anteil der Pa-

tienten mit Persönlichkeitsstörungen in einer internationalen Studie der WHO (n = 716) deutlich höher bei 39,5 % für Persönlichkeitsstörungen, die nach ICD-10 diagnostiziert wurden, und 51,1 % für Persönlichkeitsstörungen, die nach DSM-III diagnostiziert wurden (Loranger et al. 1994). 2015 wurden in Deutschland 31.409 Patienten mit der Hauptdiagnose „F60 Spezifische Persönlichkeitsstörungen“, 4950 Patienten mit der Hauptdiagnose „F61 Kombinierte und andere Persönlichkeitsstörungen“ und 611 Patienten mit der Hauptdiagnose „F21 Schizotype Störung“ vollstationär behandelt (Statistisches Bundesamt (Destatis) 2016a).

1.3 Veränderungsmessung

Verlaufs- bzw. Veränderungsmessungen erfassen Veränderungen und geben so intraindividuelle Prozesse wieder (Stieglitz und Baumann 2001). Sie eignen sich zum Nachweis der Wirksamkeit einer Therapie. Ein solcher Nachweis im Rahmen einer standardisierten Evaluation ist unter anderem notwendig, um die entstehenden Therapiekosten zu rechtfertigen (Köck und Lutz 2012).

Darüber hinaus können Verlaufsmessungen zur Erfassung des individuellen Fortschritts während einer Therapie im Rahmen der patientenorientierten Psychotherapieforschung genutzt werden. Sie ermöglichen unabhängig von der angewandten Therapiemethode die Feststellung des Fortschritts und des Ergebnisses (Köck und Lutz 2012). Die ermittelten Werte sollen zugunsten des Patienten noch während seiner Therapie zur empirisch begründeten Anpassung des Umfangs oder der Art der Therapie verwendet werden und können dadurch die Therapie optimieren (Köck und Lutz 2012; Stieglitz und Hiller 2014). Köck und Lutz (2012) führen zahlreiche Gründe für Verlaufsmessungen an: Diese können den klinischen Eindruck von einem Patienten erweitern und damit erleichtern, den Zeitpunkt für das Therapieende festzulegen oder Patienten mit schlechtem Therapieansprechen früh ausfindig zu machen. Therapieverlaufsdaten können als Anstöße in der Supervision und zur Reflexion der Behandlung durch den Therapeuten dienen. Die Bearbeitung der Verlaufsdaten mit dem Patienten kann die Motivation sowohl des Patienten als auch des Therapeuten erhöhen und deren therapeutische Beziehung verbessern und durch die strukturierte Rückmeldung an den Patienten selbst einen therapeutischen Effekt erzielen (Köck und Lutz 2012).

Die Verlaufsmessung kann sowohl qualitative als auch quantitative und reversible oder irreversible Prozesse erfassen (Stieglitz und Baumann 2001). Bei Veränderungsmessungen, bei denen der Zustand vor der Therapie als Bezugspunkt gilt, kann die Veränderung direkt (s. 1.3.1) oder indirekt (s. 1.3.2) bestimmt werden (Stieglitz und Baumann 2001). Weitere Zugangswege zur Einschätzung von Veränderungen auf der Personenebene sind die Beurteilung des Grades der Therapiezielverwirklichung und die Beurteilung des psychopathologischen Status nach Therapieende anhand von Normwerten oder Cut-off-Werten und un-

ter Berücksichtigung der klinischen Signifikanz (Seidenstücker und Baumann 1978; Stieglitz und Baumann 2001).

1.3.1 Direkte Veränderungsmessung

Die direkte Veränderungsmessung geht auf Bereiter (1963) zurück, der eine subjektive Veränderungserfassung befürwortet, bei der die Veränderung einer Variablen geradewegs eingeschätzt wird. Die Beschreibung der Veränderung wird dabei im Komparativ formuliert (Stieglitz und Baumann 2001).

Die Vorteile der direkten Veränderungsmessung werden im Folgenden vorgestellt. Das Physikalismus-Subjektivismus-Dilemma, welches auch als Bedeutsamkeitsproblem bezeichnet wird, ist gelöst, bzw. wird umgangen (Baumann et al. 1980; Bereiter 1963; Kriebel et al. 2001). Das Physikalismus-Subjektivismus-Dilemma hinterfragt, ob sich die Kontinua auf der Ebene der physikalischen Messwerte und des subjektiv-psychologischen Empfindens entsprechen und ob somit gleiche Differenzwerte an unterschiedlichen Stellen der Messwerteskala für den Patienten subjektiv gleich bedeutsam sind (Krampen und Hank 2008). Dieses Dilemma findet sich bei der indirekten Verlaufsmessung. Die gemessenen Werte der direkten Veränderungsmessung entsprechen besser der subjektiv erlebten Veränderung als indirekt ermittelte Werte (Grawe 1982).

Auch das Reliabilitäts-Validitäts-Dilemma wird umgangen (Bereiter 1963; Grawe 1982; Waldmann und Petermann 2014). Dem Dilemma liegen folgende Annahmen zugrunde: In der klassischen Testtheorie kann der wahre Merkmalswert nur von bedingungs- und zeitunabhängigen, stabilen Variablen erfasst werden. Intraindividuelle Veränderungen werden dementsprechend als Fehler gedeutet. Bei der Differenzbildung der Messwerte zweier Zeitpunkte summieren sich die Messfehler, was bedeutet, dass die Reliabilität abnimmt. Um eine hohe Reliabilität zu erzielen, dürften die Messwerte der beiden Messzeitpunkte nicht miteinander korrelieren. Eine solche niedrige Korrelation entspricht aber einer niedrigen Validität, also der Tatsache, dass unterschiedliche Inhalte gemessen werden. Folglich kann die Differenzbildung dann inhaltlich nicht mehr sinnvoll verwendet werden (Krampen und Hank 2008). Da die Abnahme der Reliabilität im Zuge einer Differenzbildung entfällt, wird erwartet, dass die Werte der direkten Veränderungsmessung reliabler sind als Differenzwerte (Grawe 1982).

Ein weiterer Vorteil der direkten Veränderungsmessung ist, dass es keine „Testdeckenprobleme“ (Grawe 1982: 240) gibt, sodass ein Patient im Rahmen der direkten Veränderungsmessung, selbst wenn er in einer vorausgehenden Testung einen maximal schlechten Wert am Ende einer Skala angegeben hat, dennoch die Möglichkeit hat, eine weitere Verschlechterung anzugeben (Grawe 1982). Das gleiche ist am anderen Ende der Skala möglich.

Darüber hinaus kommt das Problem der Regression zur Mitte (Krampen und Hank 2008; Kriebel et al. 2001; Stieglitz und Baumann 2001), wie es bei der indirekten Veränderungs-

messung besteht, nicht zum Tragen. Unter Regression zur Mitte wird das Phänomen verstanden, dass Patienten mit extremen Messwerten in der ersten Messung unabhängig von einer Therapie statistisch in der zweiten Messung einen Wert aufweisen, der näher am Durchschnittswert liegt. Dementsprechend wird ein höherer Veränderungswert erwartet und die Veränderung wird bei diesem Patienten überbewertet (Krampen und Hank 2008; Kriebel et al. 2001).

Ein weiteres Argument für die direkte Veränderungsmessung ist, dass es für die Befragten einfacher ist, Veränderungen anzugeben als Absoluturteile zu fällen (Grawe 1982). Ein intraindividueller Vergleich von Vorher zu Nachher stimmt nämlich mit der Eigenschaft eines Urteilsprozesses überein, bei dem in Relationen wahrgenommen und qualifiziert wird (Deneke 1982).

Hinzu kommt, dass die erhaltenen Messwerte leicht zu interpretieren sind (Grawe 1982) und die direkte Veränderungsmessung als ökonomischer als die indirekte Form gilt (Krampen und Hank 2008).

Den o.g. Vorteilen stehen die im Folgenden aufgeführten Nachteile bzw. kritisch zu betrachtenden Aspekte der direkten Veränderungsdiagnostik gegenüber. Zu diesen Nachteilen zählt beispielsweise, dass ohne zusätzliche Informationen nicht bekannt ist, von welchem Niveau aus die Veränderungen beginnen (Baumann et al. 1980; Grawe 1982).

Bei Mehrpunkterhebungen können sich zudem Auswertungsprobleme ergeben, bzw. können die Daten unübersichtlich erscheinen und in der Auswertung aufwendig sein (Baumann 1982; Grawe 1982).

Darüber hinaus wird kritisch angemerkt, dass die Beurteilung abhängig von der Gedächtnisleistung des Befragten ist und ein konstantes Bezugssystem voraussetzt. Teilweise werden Zustände vom Anfang nicht mehr richtig erinnert, sodass ein Problem als nicht vorhanden oder unverändert bewertet wird. Dies entspricht einer Bewertungsänderung. Lediglich das Bezugssystem zum Ende der Messung ist bekannt. Es kann sich aber im Vergleich zum früheren Zeitpunkt verändert haben (Baumann 1982; Baumann et al. 1980; Beutler und Crago 1983; Krampen und Delius 1981). Die direkte Verlaufseinschätzung wird stark beeinflusst durch die zu Therapieende noch bestehende Symptomatik und erfolgt unter zu geringer Betrachtung des Ausgangsniveaus (Schulte 2000).

Ein weiteres Argument gegen die direkte Veränderungsmessung ist, dass diese bei kurzen Zeitintervallen weniger geeignet ist (Baumann 1982), es sei denn, es werden starke Veränderungen in diesem Zeitraum erwartet (Baumann et al. 1980).

Während Baumann und Mitarbeiter (1980) herausstellen, dass bei der Anwendung der direkten Veränderungsmessung nicht häufiger Antworttendenzen in Richtung Besserung bestehen als bei der indirekten Messung, fordern Krampen und von Delius (1981) eine weitere kritische Beobachtung dieses Aspektes.

Ferner bleibt zu bedenken, dass die Urteilsbildung komplexer als eine Statusbeurteilung ist, was als Fehlerquelle wirken kann (Stieglitz und Baumann 2001) oder ihre Reliabilität senken könnte (Kriebel et al. 2001). Hinzu kommt, dass unklar ist, welche Kennwerte zur Bestimmung der Zuverlässigkeit der direkten Veränderungsmessung genutzt werden können (Grawe 1982).

1.3.2 Indirekte Veränderungsmessung

Im Rahmen der indirekten Veränderungsmessung wird der Zustand eines Patienten zu zwei Zeitpunkten mit derselben Methode ermittelt (Spitzer et al. 2012) und die Differenz der Messwerte bestimmt (Stieglitz und Baumann 2001). Es kann geprüft werden, ob die Messwertedifferenz durch Abweichungen im Rahmen der zufälligen Wahrscheinlichkeit zustande kommt oder ob eine statistisch signifikante Veränderung vorliegt (Mößner und Kordy 2014; Spitzer et al. 2012). Jacobson und Truax (1991) entwickelten den Reliable Change Index, der sich umgeformt wie folgt darstellt: $\mathrm{RC} = \frac{x_2 - x_1}{\mathrm{SD} \cdot \sqrt{2 \cdot (1-\alpha)}}$ mit x_2 als Messwert der zweiten Messung, x_1 als Messwert der ersten Messung, SD als Standardabweichung der Gruppe vor der Therapie und α als Test-Retest Reliabilität. Wenn der Wert für RC größer als 1,96 ist, ist es unwahrscheinlich ($p < 0{,}05$), dass dem zweiten Messwert keine signifikante Veränderung zugrunde liegt (Jacobson und Truax 1991). Dies erlaubt die Darstellung in der Form, dass eine Differenz zwischen Anfangs- und Endwert als statistisch signifikant gilt, wenn sie größer ist als die kritische Differenz $\mathrm{D_{krit}} = 1{,}96 \cdot \mathrm{SE_{Diff}}$ mit einem Standardfehler der individuellen Rohwertedifferenz von $\mathrm{SE_{Diff}} = \mathrm{SD} \cdot \sqrt{2 \cdot (1-\alpha)}$ (Spitzer et al. 2012). Jacobson und Truax (1991) operationalisierten zusätzlich zur statistischen Signifikanz das Kriterium der klinischen Signifikanz. Eine Veränderung gilt als klinisch signifikant, wenn durch sie ein statistisch signifikant gebesserter Patient aus einem dysfunktionalen Bereich heraus oder/und in einen gesunden, funktionalen Bereich hinein gelangt (Jacobson und Truax 1991; Mößner und Kordy 2014; Spitzer et al. 2012; Stieglitz und Baumann 2001). Wenn die Verteilungen der Gesunden und Kranken bekannt sind, sie die gleiche Varianz aufweisen und sich die Verteilungen schneiden, gilt der Schnittpunkt als der Wert, der für eine klinisch signifikante Veränderung überschritten werden muss (s. Abbildung 1) (Jacobson und Truax 1991; Spitzer et al. 2018; 2019). Er stellt bei normalverteilten Werten und gleicher Varianz die Mitte zwischen den beiden Gruppenmittelwerten dar (Mößner und Kordy 2014). Für unterschiedliche Varianzen kann die Schwelle wie folgt berechnet werden: $c = \frac{\mathrm{SD_N} * \mathrm{M_P} + \mathrm{SD_P} * \mathrm{M_N}}{\mathrm{SD_P} * \mathrm{SD_N}}$ mit M_N als Mittelwert und SD_N als Standardabweichung der Normstichprobe sowie M_P als Mittelwert und SD_P als Standardabweichung der Stichprobe vor der Behandlung (Jacobson und Truax 1991).

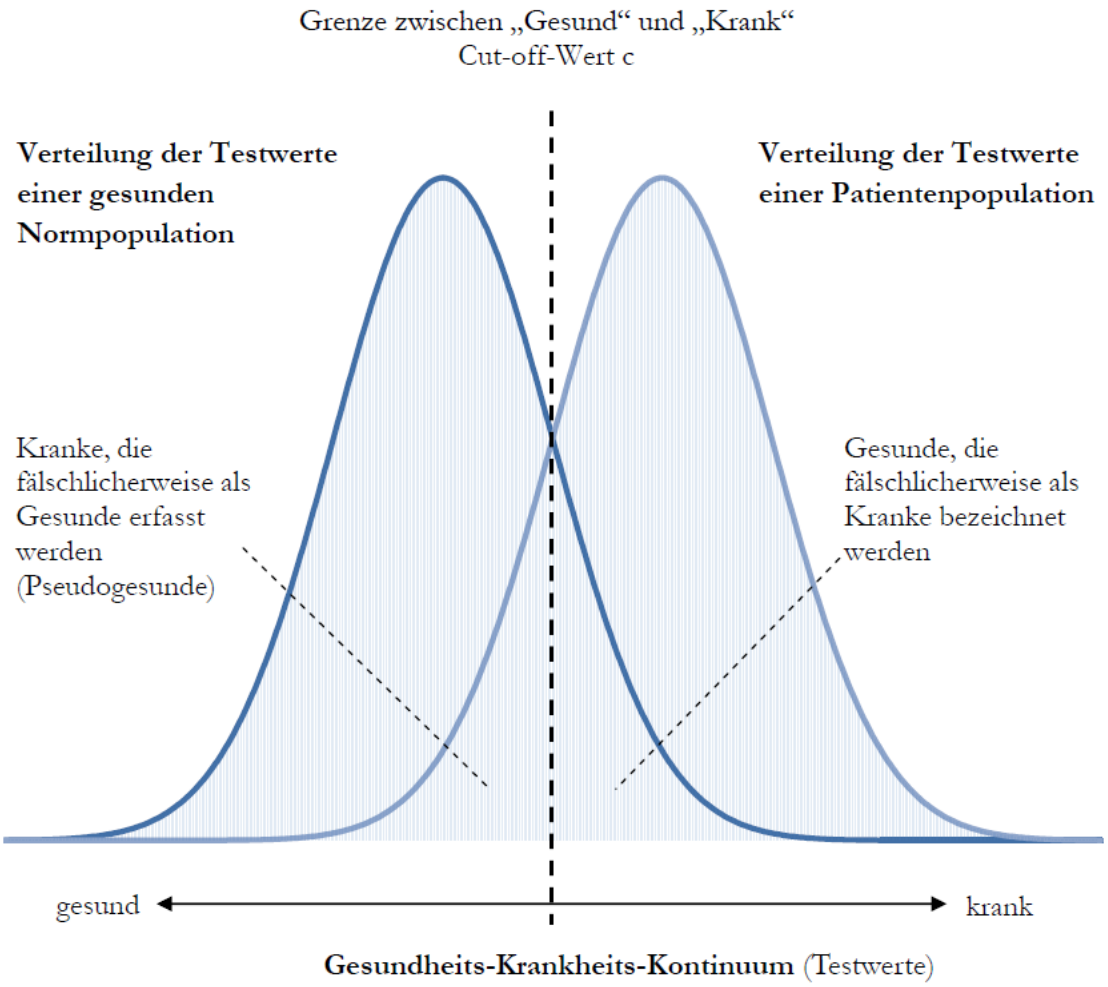

Abbildung 1: Verteilung der Gesunden und Kranken (zusammengefasst aus Jacobson und Truax 1991; Percevic et al. 2004; Spitzer et al. 2019)

Basierend auf dem zweistufigen Ansatz der statistischen und klinischen Signifikanz können folgende Therapieverläufe unterschieden werden: „Heilung: statistisch und klinisch signifikant verbessert. Besserung: nur statistisch signifikant verbessert. Nonresponse (unverändert): keine statistisch signifikante Veränderung. Verschlechterung: statistisch signifikante Verschlechterung […]. Klinisch bedeutsame Verschlechterung: statistisch und klinisch signifikant verschlechtert“ (Spitzer et al. 2012: 418). Als pseudogesund gilt, wer ungeachtet möglicher Veränderungen vor und nach der Therapie Werte erreicht, die einer gesunden Stichprobe entsprechen (Davies-Osterkamp et al. 1996; Spitzer et al. 2018; 2019). Der Therapieerfolg in der Gruppe der Pseudogesunden ist nicht auf Symptomebene erfassbar (Spitzer et al. 2012; 2018; 2019).

1.3.3 Direkte und indirekte Veränderungsmessung im Vergleich

Die direkte und indirekte Veränderungsmessung unterscheiden sich hinsichtlich der Anzahl der Messzeitpunkte, der Messwerte, des Messvorgangs und der Einsatzmöglichkeiten bei kurzem Messabstand, wie in Tabelle 5 dargestellt ist.

Tabelle 5: Unterschiede zwischen der direkten und indirekten Veränderungserfassung nach Stieglitz und Baumann (2001)

	Direkte Veränderungsmessung	Indirekte Veränderungsmessung
Anzahl der Messzeitpunkte	einer	mindestens zwei
Messwerte	direkt gemessene Veränderungswerte	abgeleitet (z. B. Differenzen)
Messvorgang	retrospektive Betrachtung von der Stärke und Richtung einer Veränderung innerhalb eines Zeitraums	mehrmalige Statuserhebungen mithilfe unabhängiger Messungen
Messabstand	weniger geeignet für kurze Zeitintervalle	bei kurzen Zeitintervallen nutzbar

Direkt und indirekt erhobene Veränderungswerte lassen sich faktorenanalytisch trennen, was bestätigt, dass es sich um zwei unterschiedliche Ansätze handelt (Schulte 2000). Stieglitz und Baumann (2001) betonen, dass „indirekte und direkte Veränderungserfassungen nicht konvergieren, sondern beide Ansätze wichtige Informationen zur Bewertung von Veränderungen liefern und daher auch beide zum Einsatz kommen sollten“ (Stieglitz und Baumann 2001: 30-31). Dass sich die beiden Ansätze ergänzen, stellen zahlreiche Autoren heraus (Baumann et al. 1980; Beutler und Crago 1983; Grawe 1982).

1.3.4 Selbst- und Fremdbeurteilungsverfahren

Abhängig von der Informationsquelle, die der Erhebung zugrunde liegt, können Selbst- und Fremdbeurteilungsverfahren unterschieden werden. Die Selbstbeurteilungsverfahren basieren auf Selbstaussagen des Patienten über das von ihm selbst beobachtete Verhalten oder sein Erleben. Die Fremdbeurteilungsverfahren nutzen darüber hinaus Fremdbeobachtungen anderer, u. a. von Ärzten, Psychologen, Pflegekräften, Angehörigen oder Freunden, sodass Informationen aus unterschiedlichen Perspektiven und Situationen einfließen (Stieglitz et al. 2001). Für beide Verfahren existieren störungsspezifische und störungsgruppenübergreifende Tests (Stieglitz et al. 2001).

Vorteile der Selbstbeurteilungsverfahren sind, dass sie zahlreiche Indikationsbereiche aufweisen (Stieglitz und Freyberger 2001), als ökonomisch gelten (Möller 1990; Stieglitz und Freyberger 2001), meist teststatistisch geprüft sind und häufig Auswertungs- und Durchführungsobjektivität aufweisen (Stieglitz und Freyberger 2001). Außerdem können manche Informationen nur durch Selbstbeurteilungsverfahren ermittelt werden und für einen Teil von Informationen ist die Selbstbeurteilung die einfachste Erhebungsart (Stieglitz und

Freyberger 2001). Die Selbstbeurteilungen können in kurzen Zeiträumen verwendet werden und auf Einzelfallebene ausgewertet werden (Möller 1990).

Ein Vorteil der Fremdbeurteilung gegenüber der Selbsteinschätzung ist, dass die Fremdbeurteilung auch zur Beschreibung komplexer Situationen, die aufgrund ihrer Komplexität das Beurteilungsvermögen eines Patienten überschreiten, genutzt werden kann (Stieglitz et al. 2001). Des Weiteren sind Fremdbeurteilungen hilfreicher bei Patienten, die nur eingeschränkt eine Selbsteinschätzung leisten können oder wollen, z. B. aufgrund ihres Krankheitsbildes (Möller 1990; Stieglitz et al. 2001). Durch die Bestimmung der Interrater-Reliabilität kann die Qualität der Fremdbeurteilung überprüft und falls nötig durch Training des Beurteilers optimiert werden (Stieglitz et al. 2001). Problematisch kann sich bei der Fremdbeurteilung die Verfügbarkeit, eine ausreichende Motivation, genügend Wissen über den Patienten, Unterschiede in der Einschätzung durch unterschiedliche Beurteiler oder ein Wechsel der Rater bei Verlaufsbeurteilungen darstellen (Stieglitz et al. 2001). Höhere Effektstärken bei der Fremdbeurteilung eines Therapieverlaufes können darauf zurückzuführen sein, dass der beurteilende Kliniker oder Forscher eventuell die Therapieergebnisse überschätzt, da er bezüglich des Therapieerfolgs positiv eingestellt ist (Stieglitz und Hiller 2014). Solche signifikant höheren Effektstärken bei Fremd- als bei Selbstbeurteilungsverfahren fanden beispielsweise Cuijpers und Kollegen (2010) in einer Metaanalyse bei der Therapie von Patienten mit Depressionen. Im Gegensatz zu unabhängigen Beurteilern überschätzen sowohl die Patienten als auch die Therapeuten den Umfang der erreichten Therapieziele (Seidenstücker und Baumann 1978).

Stieglitz und Hiller (2014) konstatieren, dass die Selbst- und die Fremdbeurteilung unterschiedliche Perspektiven auf dieselbe Erscheinung darstellen, die komplementär (Stieglitz und Hiller 2014; Möller 1990) verwendet werden. Seidenstücker und Baumann (1978) nach nutzen die Selbst- und die Fremdbeurteilung jedoch unterschiedliche Bezugspunkte zur Beurteilung, da die Fremdeinschätzung sich stärker an beobachtbarem Verhalten orientiert und bei der Selbstbeurteilung die Patienten stärker die Veränderung des Erlebens bewerten. Gemeinsam ist zahlreichen Autoren, dass sie die Verwendung mehrerer Datenquellen im Rahmen einer multimodalen Diagnostik und damit den Einsatz sowohl von Selbst- als auch von Fremdbeurteilungsverfahren fordern (Baumann und Stieglitz 2001; Seidenstücker und Baumann 1978; Stieglitz und Hiller 2014; Stieglitz et al. 2001). Selbst- und Fremdbeurteilungsverfahren gelten wissenschaftstheoretisch insgesamt als gleich gut (Baumann und Stieglitz 2001).

1.3.5 Probleme und Ansprüche bei der Veränderungsmessung

Bisher gibt es keine diagnostischen Standards zur Veränderungsmessung (Stieglitz und Hiller 2014). Probleme, die die Rahmenbedingungen der Veränderungsmessung betreffen, erstrecken sich auf Gedächtnisprozesse, Aspekte der Skalierung, das Urteilsverhalten, den Bezugspunkt der Beurteilung, Messzeitpunkte und Faktoren auf der Patientenseite selbst

(Stieglitz und Baumann 2001). Diese werden im Folgenden genauer beschrieben. Bezüglich des Gedächtnisses besteht die Problematik, dass Faktoren wie etwa die Stimmung die aktuelle Erinnerungsfähigkeit verändern können (Stieglitz und Baumann 2001). Es ist fraglich, ob die mit Begriffen umschriebenen Skalenstufen eines Tests im Zeitverlauf gleich verstanden werden und somit Differenzwerte sinnvoll interpretiert werden können (Stieglitz und Baumann 2001). Problematisch ist auch, dass die Art zu bewerten intra- und interindividuell variabel ist (Stieglitz und Baumann 2001). Zusätzlich kann sich das Bezugssystem des Patienten und damit die Gewichtung einzelner Items verändern, obwohl diese als gleich bedeutsam in die Auswertung einfließen (Stieglitz und Baumann 2001). Abhängig von der Fragestellung müssen die Zeitpunkte und die zeitlichen Abstände zwischen zwei Messungen gewählt werden (Stieglitz und Baumann 2001; Stieglitz und Hiller 2014), wovon abhängt, welche Testverfahren angewandt werden können (Stieglitz und Hiller 2014). Auf der Patientenseite können „Persönlichkeitsmerkmale (z. B. Einfluss von Neurotizismus auf die Erinnerungsleistung), klinische Merkmale (z. B. Erwartungshaltungen, Motivation) oder auch Urteilstendenzen (z. B. Simulation/Dissimulation, Aggravation)“ (Stieglitz und Baumann 2001: 22) die Veränderungsmessung beeinflussen. Zusätzlich sehen Stieglitz und Hiller (2014) aufseiten der Anwender das Problem, dass diese bisher selten Veränderungsmessungen anwenden und wenig über die Einsatzmöglichkeiten wissen.

Neben den Problemen bezüglich der Rahmenbedingungen bestehen messtheoretische Schwierigkeiten und Defizite bei der Testkonstruktion und -entwicklung (Stieglitz und Baumann 2001; Stieglitz und Hiller 2014). In Hinblick auf die Messtheorie kritisieren Stieglitz und Hiller (2014), dass bei der Entwicklung von Tests diese nach der Klassischen Testtheorie konzipiert werden, bei der eine Veränderungsmessung nicht vorgesehen ist. Bisher wurde der Entwicklung von Veränderungsmaßen wenig Aufmerksamkeit geschenkt (Ogles 2013), sodass zurzeit häufig Tests zur Statusdiagnostik zur Veränderungsmessung eingesetzt werden (Stieglitz und Baumann 2001). Dabei ergeben sich Schwierigkeiten bei der Bewertung von Veränderungen. Eine Schwierigkeit ist, dass die Verteilungen der klinischen Population und der Referenzpopulation bekannt und die Skalenwerte in beiden Populationen normalverteilt sein müssen, um nach der Berechnung eines Trennwertes klinisch relevante Veränderung erkennen zu können. Hierzu fehlen jedoch oft verlässliche Daten (Mößner und Kordy 2014; Stieglitz und Hiller 2014). Auch liegen bisher wenig metaanalytische Daten zu Cut-off-Werten und Reliable Change Indizes einzelner Tests vor, sodass klinisch signifikante Veränderungen in verschiedenen Studien unterschiedlich bestimmt werden (Stieglitz und Hiller 2014). Mößner und Kordy (2014) bezeichnen die klinische Signifikanz sogar als relatives Konzept, da die Zuweisung einer Person zur gesunden oder klinischen Gruppe von der Entfernung zwischen der gesunden und der klinischen Population abhängig ist. Wenn die Personen der klinischen Referenzpopulation im Mittel besonders hohe Werte auf einer Skala aufweisen, ergibt sich nämlich ein höherer Trennwert, als wenn die Referenzpopulation weniger hohe Werte im Mittel zeigt. Dadurch kann es vorkommen, dass eine Person mit einem bestimmten Skalenwert in den unterschiedlichen

Fällen einmal in der gesunden und einmal in der klinischen Population eingeordnet wird. Umgangen wird dieses Problem, indem der Trennwert nur in Bezug auf eine feste Referenzpopulation festgelegt wird (Mößner und Kordy 2014).

Abgeleitet von den genannten Problemen werden spezielle Anforderungen an Instrumente zur Veränderungsmessung gestellt. Beispielsweise verlangen Stieglitz und Hiller (2014) eine Berücksichtigung des Verlaufsaspektes schon bei der Itemauswahl und bei der psychometrischen Analyse eines Tests. Für Tests, die zur Veränderungsmessung genutzt werden, wird gefordert, dass Testkennwerte wie die interne Konsistenz, die Paralleltest-Reliabilität pro Messzeitpunkt, die Trennschärfe und, falls der Test auf einer Faktorenanalyse beruht, die Faktorenstruktur stabil sind (Stieglitz et al. 1980). Außerdem sollten sie auf Änderungssensitivität geprüft sein. Änderungssensitivität kann in der Testkonstruktion durch Verwendung änderungssensitiver Items erreicht werden, also durch den Einsatz von Items, die Veränderungen sensitiv abbilden, indem sich ihre statistischen Kennwerte durch eine Intervention verändern (Bereiter 1963; Stieglitz und Hiller 2014). Dazu gehören u. a. Items, mit großer absoluter Veränderung und solche, die stark mit dem Gesamtwert der Veränderungen korrelieren (Stieglitz und Hiller 2014). Alternativ können Skalen als Ganzes auf Änderungssensitivität im Prä-post-Vergleich geprüft werden und darauf, inwiefern sie Unterschiede zwischen einer unbehandelten und einer behandelten Gruppe erkennen (Stieglitz und Hiller 2014). Neben der Änderungssensitivität gehören die Assoziation mit dem Globalurteil, die Unterscheidung zwischen wirksamer und unwirksamer Therapie und die Übereinstimmung empirischer Daten, die mittels einer Ratingskala erfasst wurden, mit den Modellannahmen des dynamischen Raschmodells zu möglichen Validitätskriterien für die Veränderungsmessung (Maier et al. 1990). Ein zusätzlich häufig geforderter Anspruch an die Veränderungsmessung ist eine multimodale Diagnostik im Zeitverlauf (s. 1.4) (Baumann und Stieglitz 2001; Seidenstücker und Baumann 1978; Stieglitz und Baumann 2001; Stieglitz und Hiller 2014).

1.4 Multimodale Diagnostik

Der Ansatz der multimodalen bzw. multimethodalen, sogenannten Mehrebenendiagnostik berücksichtigt zur Diagnostik nicht nur eine, sondern mehrere bewusst ausgesuchte Messmethoden (Baumann und Stieglitz 2001). In der Praxis dient dies der Verbesserung der Entscheidungssicherheit (Seidenstücker und Baumann 1978). In der multimethodalen Diagnostik werden Variationen bezüglich der Datenebenen (Grundkategorien organismischer Merkmale, z. B. Leistungen, Erleben, Verhalten), der Datenquellen (Informationsgeber, z. B. Patient, Therapeut, Instrument), der Funktionsbereiche (Klassen psychischer Funktionen z. B. Informationsaufnahme wie Wahrnehmung, Informationsverarbeitung wie Denken, biologische Funktionen wie vegetative Veränderungen oder Reaktionsklassen wie selbstbewusstes Verhalten, Emotionen, Motorik) und der Untersuchungsverfahren (z. B. Selbst- und Fremdbeurteilungsverfahren, Verhaltensbeobachtungen, Interviews) eingesetzt

(Baumann und Stieglitz 2001; Seidenstücker und Baumann 1978). Zur Auswahl der diagnostischen Verfahren müssen die Objektivität, Reliabilität, Validität, Sensitivität, das Kosten-Nutzen Verhältnis, die Zumutbarkeit, soziale und ethische Aspekte und die Durchführbarkeit der Tests berücksichtigt werden (Baumann und Stieglitz 2001). Auch müssen die angewandten Konstrukte und Operationalisierungen der Funktionsbereiche, die bei dem Patienten verändert sind, beachtet werden (Seidenstücker und Baumann 1978). Durch den Einsatz multipler Verfahren können sich Interpretationsschwierigkeiten ergeben, wenn die einzelnen Ergebnisse nicht übereinstimmen (Baumann und Stieglitz 2001; Möller 1990). Falls die Ergebnisse von Verfahren, die zum gleichen Zeitpunkt durchgeführt wurden, sich unterscheiden, wird von Diskordanz als Gegensatz zur Konkordanz gesprochen. Synchronizität und Desynchronizität beschreiben den Grad der Übereinstimmung von Untersuchungsergebnissen zu unterschiedlichen Messzeitpunkten (Baumann und Stieglitz 2001; Möller 1990).

Speziell in der Misserfolgsforschung wird ein multimodales Vorgehen gefordert: „Nicht unproblematisch ist [...], wenn der Behandlungserfolg anhand nur eines einzigen Outcomekriteriums, zum Beispiel der Veränderung der Symptombelastung, beurteilt wird, anstatt anhand eines breitgefächerten Spektrums von Kriterien, die möglichst unterschiedliche Perspektiven der Evaluation (Patient, Therapeut, Angehörige) einbeziehen“ (Reuter et al. 2014: 136). „Die Gefahr von Zufallsentscheidungen [kann] nur durch multiple (Miss-) Erfolgskriterien reduziert werden“ (Spitzer et al. 2012: 422), da der Grad der Übereinstimmung der einzelnen Messmethoden als unbefriedigend zu bewerten ist (Spitzer et al. 2012).

1.5 Zusammenfassung und eigene Fragestellung

Im Folgenden wird der aktuelle Stand der Forschung mit seinen Limitationen zusammengefasst und die hiervon abgeleitete Fragestellung dieser Arbeit vorgestellt.

1.5.1 Zusammenfassung und Bewertung der aktuellen Forschungslage

Wie in den vorausgehenden Kapiteln dargestellt wurde, ist das Auftreten von Nonresponse und Verschlechterungen in der stationären Psychotherapie bisher noch nicht umfassend untersucht worden. Die wenigen Studien weisen zudem teilweise methodische Schwachstellen auf und haben keine einheitlichen Ergebnisse hervorgebracht. Die Häufigkeitsangaben für Nonresponse unterscheiden sich um den Faktor 9 (4,9 % (Junge und Ahrens 1996) bis 44,3 % (Geiser et al. 2001)), die Angaben für Misserfolg um den Faktor 10 (1,8 % (Junge und Ahrens 1996) bis 18,8 % (Sammet et al. 2004)). Da erfolglose Therapien weitreichende Konsequenzen für den Patienten, sein Umfeld, das Behandlungsteam und das Gesundheitssystem haben (Reuter et al. 2014), ist eine multimodale Erfassung von Nonresponse und Verschlechterungen auf Einzelfallebene, die über die Bestimmung von Effektstärken hinausgeht, relevant, jedoch bisher noch nicht umfassend erfolgt (s. 1.4). Die Suche nach Patientenmerkmalen, die einen negativen Therapieverlauf erwarten lassen, ist

naheliegend. Speziell die Frage, ob Männer und Frauen sich in der Häufigkeit von Nonresponse und Verschlechterung unterscheiden, wurde in der Literatur zwar schon wiederholt untersucht, aber nicht einheitlich beantwortet (s. 1.2.3). Den bisherigen Studien zu Nonresponse und Verschlechterungen unter Psychotherapie liegen meist nur sehr kleine Stichproben zugrunde (Franz et al. 2000; Geiser et al. 2001; Geyer et al. 2003) oder sie kommen der Forderung nach multimodaler Diagnostik (Stieglitz und Baumann 2001; Stieglitz und Hiller 2014; Ogles 2013) nicht nach. Obwohl 2015 in deutschen Krankenhäusern 36.970 Patienten mit Persönlichkeitsstörungen als Hauptdiagnose (Statistisches Bundesamt (Destatis) 2016a) behandelt wurden und Persönlichkeitsstörungen somit in der stationären Psychotherapie ein relevantes Krankheitsbild darstellen, existiert nur die Studie von Spitzer und Kollegen (2012) zu Misserfolgen in der stationären Psychotherapie von Patienten mit Persönlichkeitsstörungen. Diese untersucht bereits eine große Stichprobe (n = 1239) unter Beachtung der unterschiedlichen Arten der Veränderungsmessung (s. 1.3) (Spitzer et al. 2012). Da jedoch als Kriterium für Misserfolg ausschließlich die allgemeine Symptombelastung herangezogen wurde (Spitzer et al. 2012), ist noch nicht bekannt, inwiefern Misserfolge hinsichtlich interpersonaler Probleme in der stationären Psychotherapie vorliegen.

1.5.2 Fragestellungen und Ziele dieser Arbeit

Anders als in der bestehenden Literatur zu Misserfolgen und Nonresponse in der stationären Psychotherapie soll eine größere Stichprobe eines Fachkrankenhauses untersucht werden. In der Analyse sollen dabei mehrere Outcome-Maße verwendet werden, indem die Symptombelastung, die Stärke interpersonaler Probleme und die Beeinträchtigungsschwere im Verlauf der Therapie für die untersuchten Gruppen betrachtet werden. Die folgenden Fragen sollen bearbeitet werden:

1. Wie häufig verschlechtern sich Patienten im Rahmen einer stationären psychotherapeutischen Behandlung und wie häufig tritt Nonresponse auf?
2. Unterscheiden sich Männer und Frauen in der Auftretenshäufigkeit von Erfolg und Misserfolg (Nonresponse und Verschlechterung) oder in der Stärke der Symptombelastungen, interpersonalen Probleme und in der Beeinträchtigungsschwere?
3. Unterscheiden sich Patienten ohne Persönlichkeitsstörungen und Patienten mit Persönlichkeitsstörungen in der Auftretenshäufigkeit von Erfolg, Nonresponse und Verschlechterung? Gibt es diesbezüglich Unterschiede zwischen den Patienten mit Persönlichkeitsstörungscluster B und C? Unterscheiden sich darüber hinaus Personen mit und ohne Persönlichkeitsstörungen sowie die Personen der Persönlichkeitsstörungscluster B und C in ihrer Symptombelastung, ihren interpersonalen Problemen und ihrer Beeinträchtigungsschwere im Verlauf der Therapie voneinander?

2 Patienten, Material und Methoden

Im Folgenden werden der Ablauf, die Ein- und Ausschlusskriterien der Studie, die statistischen Verfahren und die verwendeten Untersuchungsinstrumente vorgestellt.

2.1 Ablauf der Studie

Es wurde eine retrospektive Datenanalyse einer beobachtenden, d.h. nicht-experimentellen und nicht-interventionellen klinischen Studie durchgeführt. Die Daten für diese naturalistische Psychotherapie-Studie wurden im Zeitraum vom 01.01.2005 bis zum 31.12.2014 im Asklepios Fachklinikum Tiefenbrunn sowohl im Bereich für Psychiatrie und Psychotherapie als auch im Bereich für Psychosomatik und Psychotherapie erhoben. Bei vorliegender Einverständniserklärung wurden alle Patienten bzw. ihre behandelnden Therapeuten darum gebeten, die in Kapitel 2.3 dargestellten Fragebögen und die Fragen des Monitoring-Programms zum Aufnahme- und Entlassungszeitpunkt zu bearbeiten. Die Patienten erhielten eine psychodynamisch orientierte stationäre Behandlung, in der sie dreimal pro Woche an einer 60- oder 90-minütigen Gruppenpsychotherapie sowie ein- bis zweimal wöchentlich an Einzelgesprächen teilnahmen. Abhängig von der Indikation wurden zusätzlich Ergo-, Gestaltungs- oder Physiotherapie und Entspannungs- oder körpertherapeutische Verfahren angewandt. Bei gegebener Indikation wurden die Patienten zusätzlich pharmakotherapeutisch behandelt. Die in Tiefenbrunn entwickelte psychoanalytisch-interaktionelle Therapie (Heigl-Evers et al. 1994; Streeck und Leichsenring 2011), bei der der Schwerpunkt der Therapie auf den interpersonellen Beziehungen des Patienten liegt (Streeck und Leichsenring 2011), wurde bei Patienten mit strukturellen Störungen eingesetzt. Dazu gehören Personen, die nicht ausreichend über die psychischen Funktionen verfügen, „die für die Organisation des Selbst und seine Beziehungen zu den inneren und äußeren Objekten erforderlich sind“ (Rudolf 2013: 54) oder bei denen diese Funktionen vulnerabel sind (Rudolf 2013). Zu diesen Personen gehören u. a. Patienten mit Persönlichkeitsstörungen (Rudolf 2013).

Durch die Mitarbeiter des Asklepios Fachklinikums Tiefenbrunn erfolgte vor der statistischen Auswertung eine Datenbearbeitung und Anonymisierung „an der Quelle“. Von Seiten der Ethik-Kommission der Universitätsmedizin Göttingen bestanden bezüglich der Studie keine Bedenken (Antrags-Nummer DOK_184_2015).

2.2 Ein- und Ausschlusskriterien

In die Studie wurden Patienten eingeschlossen, die im Zeitraum vom 01.01.2005 bis zum 31.12.2014 mindestens sechs Wochen im Fachklinikum Tiefenbrunn behandelt wurden, die mindestens 18 Jahre alt waren und für die der Global Severity Index der SCL-90-R als Ergebnismaß sowohl bei Aufnahme als auch bei Entlassung erfasst wurde. Diese Einschluss-

kriterien erfüllten 3862 (59,6 %) der 6481 in diesem Zeitraum in Tiefenbrunn behandelten Patienten.

2.3 Untersuchungsinstrumente

Bei dieser Studie wurden drei Untersuchungsinstrumente im Sinne der indirekten Veränderungsmessung und zwei Verfahren zur direkten Veränderungsmessung eingesetzt. Für beide Arten der Veränderungsmessung wurden hierbei Instrumente sowohl zur Selbst- als auch zur Fremdbeurteilung verwendet (s. Tabelle 6).

Tabelle 6: Übersicht über die verwendeten Erhebungsinstrumente

	Selbstbeurteilungs-verfahren	Fremdbeurteilungs-verfahren
Indirekte Veränderungsmessung	SCL-90-R (s. 2.3.1) IIP (s. 2.3.2)	BSS (s. 2.3.3)
Direkte Veränderungsmessung	Ratingskala zur Einschätzung der Veränderung der Symptomatik (s. 2.3.4)	Ratingskala zur Einschätzung der Veränderung der Symptomatik (s. 2.3.4)

Die Patienten schätzten zu Beginn der Therapie und bei der Entlassung ihre körperliche und psychische Symptombelastung unter Verwendung der Symptom-Checkliste SCL-90-R und ihre interpersonalen Probleme unter Verwendung des Inventars zur Erfassung interpersonaler Probleme IIP ein. Die behandelnden Therapeuten beurteilten zu den gleichen Zeitpunkten mithilfe des Beeinträchtigungs-Schwere-Scores BSS die Schwere der Erkrankung ihrer Patienten. Sowohl die Patienten als auch die Therapeuten schätzten darüber hinaus im Rahmen der direkten Veränderungsmessung nach der Therapie die Veränderung der Symptomatik des Patienten ein. Die verwendeten Fragebögen bzw. Testverfahren werden in den Kapiteln 2.3.1 bis 2.3.4 vorgestellt. Zur Beurteilung der Tests werden jeweils Angaben zu den Hauptgütekriterien Objektivität, Reliabilität und Validität gemacht.

Die Objektivität eines Tests gibt an, inwieweit ein Testergebnis unabhängig von dem Untersucher ist. Man unterscheidet dabei die Durchführungs- von der Auswertungs- und der Interpretationsobjektivität, abhängig davon, zu welchem Zeitpunkt im Testverfahren die Objektivität betrachtet wird (Lienert und Raatz 1998).

Die Reliabilität wird definiert als Grad der Genauigkeit, mit dem ein Test ein bestimmtes Merkmal misst, unabhängig davon, ob es das Merkmal ist, welches der Test beabsichtigt zu messen (Schmidt-Atzert und Amelang 2012; Lienert und Raatz 1998). Die Reliabilität entspricht dem Anteil der wahren, messfehlerfreien Varianz an der gesamten Varianz (Bühner 2011; Lienert und Raatz 1998), sodass der Reliabilitätskoeffizient Werte zwischen null und eins annehmen kann (Rost 2004; Schmidt-Atzert und Amelang 2012). Zu den klassischen

Methoden der Reliabilitätsberechnung gehören die Retest- und die Paralleltestmethode sowie die Bestimmung der internen Konsistenz eines Tests mittels der Testhalbierungsmethode oder einer Konsistenzanalyse (Lienert und Raatz 1998). Bei der Retestmethode wird derselbe Test an derselben Stichprobe nach einer ausreichend langen Zeit wiederholt. Der Korrelationskoeffizient, der sich aus der Korrelation der Testergebnisse der beiden Testdurchführungen berechnen lässt, entspricht der Retest-Reliabilität (Lienert und Raatz 1998; Schmidt-Atzert und Amelang 2012). Werden stattdessen zwei parallele Testformen in einem kurzen Zeitintervall an derselben Stichprobe durchgeführt, entspricht die Korrelation der Ergebnisse der beiden Tests der Paralleltestreliabilität (Lienert und Raatz 1998). Bei der Testhalbierungsmethode wird die Testhalbierungsreliabilität ermittelt, indem der Test einmal an einer Stichprobe durchgeführt wird und die Rohwerte einer Testhälfte mit den Werten der zweiten, äquivalenten Hälfte korreliert werden (Lienert und Raatz 1998). Wird der Test nicht nur in zwei Hälften, sondern in mehrere äquivalente Teile geteilt, bis hin zu einer Einteilung in so viele Abschnitte, wie Items vorliegen, kann die interne Konsistenz mithilfe unterschiedlicher Formeln berechnet werden (Lienert und Raatz 1998). Bei gleich langen äquivalenten Testteilen und bei intervallskalierten Antwortmöglichkeiten kann beispielsweise Cronbachs α-Koeffizient zur Abschätzung der Reliabilität herangezogen werden (Lienert und Raatz 1998).

Die Validität gibt an, in welchem Ausmaß ein Test dasjenige Merkmal wirklich misst, welches er messen soll (Lienert und Raatz 1998; Rost 2004; Schmidt-Atzert und Amelang 2012). Es können unterschiedliche Aspekte der Validität unterschieden werden. Die Inhaltsvalidität, die angibt, ob ein Test das zu messende Merkmal optimal abbilden kann, wird einem Test meist durch Expertenbeurteilungen zugeschrieben (Lienert und Raatz 1998). Für die Kriteriumsvalidität werden die Testergebnisse einer Stichprobe mit einem konkreten, messbaren Außenkriterium korreliert (Lienert und Raatz 1998; Schmidt-Atzert und Amelang 2012). Abhängig vom zeitlichen Verhältnis zwischen der Durchführung des Tests und der Erfassung des Kriteriums wird Übereinstimmungs- oder Vorhersagevalidität als Kriteriumsvalidität angenommen (Schmidt-Atzert und Amelang 2012). Die Konstruktvalidität gibt Auskunft darüber, inwiefern ein Test ein bestimmtes Konstrukt messen kann (Lienert und Raatz 1998; Schmidt-Atzert und Amelang 2012). Um zu zeigen, dass ein Test ein bestimmtes Konstrukt besser misst als ein anderes Konstrukt, wird einerseits die konvergente Validität bestimmt, die den Zusammenhang mit anderen Aspekten des Konstrukts untersucht, und andererseits die divergente bzw. diskriminante Validität ermittelt, die den Zusammenhang mit Aspekten anderer Konstrukte beurteilt (Schmidt-Atzert und Amelang 2012). Für eine hohe Konstruktvalidität muss die konvergente Validität dementsprechend deutlich höher sein als die diskriminante Validität (Schmidt-Atzert und Amelang 2012). Darüber hinaus gehört die faktorielle Validität zur Konstruktvalidität. Diese entspricht der faktoriellen Struktur eines Tests, die sich im Rahmen einer Faktorenanalyse des Tests mit einigen Außenkriterien und konstruktnahen und -fernen Tests ergibt (Lienert und Raatz 1998).

2.3.1 Die Symptom-Checkliste SCL-90-R von L. R. Derogatis

Die Symptom-Checkliste gehört zu den störungsgruppenübergreifenden Selbstbeurteilungsverfahren (Stieglitz und Freyberger 2001). Sie ermöglicht die Beurteilung der subjektiven Beeinträchtigung durch körperliche und psychische Symptome innerhalb der letzten sieben Tage. Dazu umfasst sie 90 Items, die auf einer fünfstufigen Likert-Skala von 0: „überhaupt nicht“ bis 4: „sehr stark“ bewertet werden. Die SCL-90-R kann bei Jugendlichen ab 12 Jahren und Erwachsenen eingesetzt werden. Eine Verwendung zur Verlaufsbeurteilung ist als Prä-Post-Messung oder als wöchentliche Testung während einer Behandlung möglich (Franke 2002).

Die Auswertung erfolgt mehrdimensional. Einzeln betrachtet können die Items Auskunft über die Belastung durch spezifische Symptome geben. 83 der 90 Items können darüber hinaus den neun Skalen Somatisierung (SOMA), Zwanghaftigkeit (ZWAN), Unsicherheit im Sozialkontakt (UNSI), Depressivität (DEPR), Ängstlichkeit (ANGS), Aggressivität/Feindseligkeit (AGGR), Phobische Angst (PHOB), Paranoides Denken (PARA) und Psychotizismus (PSYC) zugeordnet werden. Diese erlauben Aussagen über die Symptombelastung in syndromalen Bereichen. Sieben Zusatzfragen, die keiner Skala zugeordnet werden, fließen ergänzend bei der Berechnung der globalen Kennwerte ein. Es können drei globale Kennwerte unter der Berücksichtigung aller 90 Items berechnet werden: der GSI (global severity index) beschreibt die generelle psychische Belastung, der PSDI (positive symptom distress index) misst die Stärke der Antworten und der PST (positive symptom total) nennt die Anzahl der Symptome, bei denen eine Belastung beschrieben wird (Franke 2002). Die Flächentransformation der Rohwerte zu T-Werten bietet eine vierte Beurteilungsebene, nämlich die Bewertung der Abweichung des Einzelfalls von einzelnen Vergleichsstichproben. Zur Beurteilung liegen die repräsentative Eichstichprobe Erwachsener (N1, n = 2141), die Normstichprobe Jugendlicher (N2, n = 857), die Normstichprobe Studierender (N3, n = 800), die Untersuchungsstichprobe Medizinstudierender (U2, n = 250) und Krankenpflegeschüler (U3, n = 98), die Stichprobe der stationären Psychotherapiepatienten (P1, n = 5057), der HIV-infizierten Patienten (P2, n = 568), der schizophrenen Psychotherapiepatienten (P3, n = 110) oder der sehbeeinträchtigten Patienten (P4, n = 117) vor (Franke 2002). Als „Fall“ bzw. als psychisch belastet gelten diejenigen, die einen T-Wert des GSI und/oder einen T-Wert von mindestens zwei Skalen von größer oder gleich 63 aufweisen (Franke 2002). Unabhängig von der Falldefinition im ersten Schritt der Auswertung der SCL-90-R empfiehlt Franke (2002) als zweiten und dritten Schritt die Untersuchung, bei welchen Globalwerten und bei welchen Skalen T-Werte größer als 60 vorkommen. Im Anschluss daran soll analysiert werden, bei welchen Items der Skalen mit $T > 60$ ein Wert $> M + 1$ SD vorliegt (Schritt 4). Auch bei der Interpretation der Zusatzitems (Schritt 5) gilt ein Wert $> M + 1$ SD als auffällig (Franke 2002). Diese fünfschrittige Vorgehensweise dient der Gesamtinterpretation der SCL-90-R (Franke 2002).

Im Manual zur SCL-90-R (Franke 2002) werden die folgenden testtheoretischen Interpretationsvorschläge bei der Veränderungsmessung angegeben: Bezüglich der indirekten Verlaufsmessung für stationäre Psychotherapiepatienten kann davon ausgegangen werden, dass bei einer Veränderung des GSI von mindestens ± 0,3 bzw. von ± 4,80 für T-Werte eine statistisch signifikante Verbesserung oder Verschlechterung vorliegt. Für eine klinisch signifikante Verbesserung muss der GSI ≤ 0,7 betragen (Franke 2002).

Die englischsprachige SCL-90-R – hervorgegangen aus der 1973 von einer Arbeitsgruppe um Derogatis veröffentlichten SCL-90 – wurde in den siebziger Jahren, die deutschsprachige Version 1986 von Derogatis publiziert (Franke 2002; 2014). 1995 wurde eine erste für den deutschsprachigen Raum gültige Normierung veröffentlicht (Franke 2002; 2014). Neue Normierungen wurden in der zweiten überarbeiteten und neu normierten Auflage 2002 von Franke (2002) vorgestellt. Diese wurden in dieser Arbeit verwendet, da es sich zum Zeitpunkt der Datenerhebung um die aktuelle Testversion handelte. 2013 erschien die neu normierte SCL-90-S, die u.a. Veränderungen mit dem Ziel einer einfacheren Sprache aufweist (Franke 2014).

Die SCL-90-R gilt international als objektives, reliables und valides Instrument (Franke 2003). Sie weist Durchführungs-, Auswertungs- und Interpretationsobjektivität auf (Franke 2003). Die Reliabilität der SCL-90-R wird anhand der internen Konsistenz und der Retest-Reliabilität beurteilt (Franke 2002). In der Eichstichprobe N1 nimmt die interne Konsistenz der Skalen Werte von $\alpha = 0{,}75$ (PHOB) bis $\alpha = 0{,}87$ (DEPR) an. In der Stichprobe stationärer Psychotherapiepatienten P1 liegen Cronbachs α Werte von $\alpha = 0{,}74$ (AGGR) bis $\alpha = 0{,}88$ (DEPR) vor. Die interne Konsistenz des GSI ist für die Normstichprobe N1 und die Stichprobe der stationären Psychotherapiepatienten P1 mit $\alpha = 0{,}97$ sehr hoch (Franke 2002). Franke (2002) bewertet die Reliabilität der SCL-90-R insgesamt als gut. In dem Manual zu der neu normierten SCL-90-S differenziert sie anhand der Werte der internen Konsistenz zwischen zufriedenstellender Reliabilität der Skala AGGR, guter Reliabilität der Skalen PARA, PHOB, PSYC, SOMA, UNSI und sehr guter Reliabilität für ZWAN, ANGS, DEPR und den GSI (Franke 2014). Die Test-Retest-Reliabilität wird als gut eingeschätzt, bei gleichzeitig bestehenden Interpretationsschwierigkeiten der Retest-Reliabilität, die sich bei der Messung von zeitlich instabilen Statemerkmalen ergeben (Franke 2002). Der SCL-90-R wird Inhaltsvalidität zugesprochen (Franke 2002). Zahlreiche Studien untersuchen die noch umstrittene faktorielle Validität (Franke 2002) sowie die konvergente und diskriminante Validität, welche Franke (2014) als befriedigend bis gut beschreibt.

2.3.2 Inventar zur Erfassung Interpersonaler Probleme IIP

Das Inventar zur Erfassung interpersonaler Probleme ist ein Fragebogen zur Selbsteinschätzung interpersonaler Probleme (Horowitz et al. 2000). Es gibt Auskunft darüber, „unter welchen interpersonalen Schwierigkeiten eine Person in welchem Maße leidet“ (Horowitz et al. 2000: 25). Das Inventar zur Erfassung interpersonaler Probleme kann bei Er-

wachsenen zur Status- und Prozessdiagnostik in der Psychotherapie eingesetzt werden (Horowitz et al. 2000). Die Langform umfasst 127 Items, die Kurzfassung IIP-C, die in dieser Studie verwendet wurde, beinhaltet 64 Items (Horowitz et al. 2000). Die Fragen beginnen im ersten Teil mit „Es fällt mir schwer..." (Horowitz et al. 1994: 2). Der zweite Teil fragt nach Aspekten, die „man im Übermaß tu[t]" (Horowitz et al. 1994: 4). Die Fragen können auf einer fünfstufigen Likert-Skala von 0 (= nicht) bis 4 (= sehr) beantwortet werden. Die 64 Items, die sowohl in der Lang- als auch in der Kurzform vorkommen, lassen sich zu acht Skalen zusammenfassen, die den Oktanten des Circumplexmodells entsprechen (s. Abbildung 2) (Horowitz et al. 2000).

Bei der Entwicklung des IIP dienten das von Leary 1957 entwickelte Circumplex-Modell und die Theorie der reziproken Emotion von Sullivan als Grundlage. Das Circumplex-Modell von Leary ist ein Kreismodell, welches im zweidimensionalen Raum die Dimensionen Zuneigung und Dominanz aufweist, zwischen denen das interpersonale Verhalten einer Person eingeordnet werden kann. Es wird angenommen, dass in interpersonalen Situationen reziproke Prozesse ablaufen, also die beteiligten Personen ihr Verhalten gegenseitig beeinflussen. Dabei verhalten sich Personen hinsichtlich der Dimension der Zuneigung ähnlich und hinsichtlich der Dominanz gegensätzlich. Dieser Theorie folgend entstehen Spannungen bei nichtkomplementärem Verhalten der beteiligten Personen (Horowitz et al. 2000). Auf der Grundlage von Sullivans Theorie der reziproken Emotion und Learys Circumplexmodell interpersonalen Verhaltens veröffentlichten Horowitz und Mitarbeiter 1988 das Inventory of Interpersonal Problems. Bereits 1987 übertrugen Strauß und Kordy das IIP ins Deutsche (Horowitz et al. 2000). Die zweite Auflage wurde 2000 als überarbeitete und an einer repräsentativen Stichprobe (n = 3047) neu normierte Auflage herausgegeben (Horowitz et al. 2000).

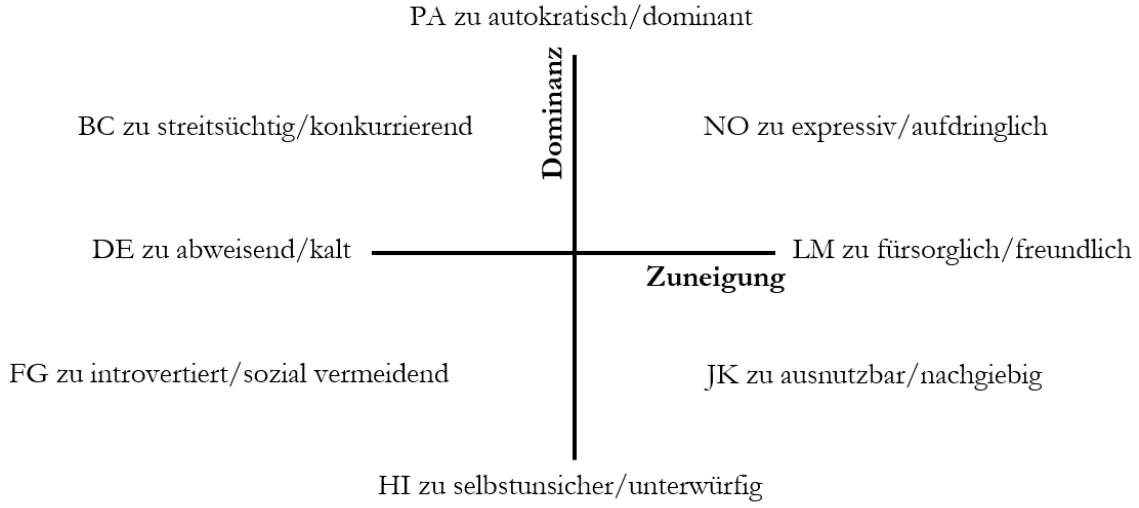

Abbildung 2: Circumplexskalen des IIP

Entsprechend des Circumplex-Modells (s. Abbildung 2) ist die Interkorrelation der im Kreismodell benachbarten Skalen am größten. Sie nimmt dann zur Skala des gegenüberlie-

genden Oktanten ab, um dem interpersonalen Zirkel folgend wieder anzusteigen (Horowitz et al. 2000).

Zur Auswertung werden die Summenwerte der einzelnen Skalen berechnet sowie der 64-Item-Gesamtwert (IIP_{ges}) als gemittelte Summe der Skalenwerte. Zum Teil werden ipsatierte Werte – Skalenwerte, von denen der IIP-64-Gesamtwert subtrahiert wurde – verwendet (Horowitz et al. 2000). Diese personenbezogene Transformation zielt darauf, „eine ‚persönliche Antworttendenz' (z. B. im Sinne einer Klagsamkeit) einer Person zu neutralisieren und so das individuelle Muster der Itemantworten bzw. ihrer Relationen stärker hervortreten zu lassen" (Horowitz et al. 2000: 13). Der Vergleich der Werte einer Person mit denen einer deutschen repräsentativen Stichprobe, die 1994 an n = 3047 Personen durchgeführt wurde, erfolgt mit Hilfe der Transformation der Werte in Stanine. Die Stanine-Skala ist eine neunstufige (1-9) standardisierte Normierungsskala, bei der der Mittelwert bei M = 5 liegt, mit einer Standardabweichung SD = 2 (Döring und Bortz 2016a). Die Stanine S_v lassen sich nach der Formel $\mathrm{S_v} = 5 + 2 * \mathrm{z_v}$ mit $\mathrm{z_v} = \frac{x_v - \bar{x}}{SD(x)}$ berechnen, wobei z_v beschreibt, um wie viele Einheiten der Standardabweichung SD(x) der Testwert x_v einer Testperson v vom arithmetischen Mittelwert $\bar{x}$ der Verteilung der Bezugsgruppe abweicht (Goldhammer und Hartig 2012). Jedoch werden nur Werte von 1 bis 9 angegeben (Schmidt-Atzert und Amelang 2012). Für das Inventar zur Erfassung interpersonaler Probleme liegen nach Alter und Geschlecht getrennte Normtabellen für ipsatierte und nicht-ipsatierte Werte vor. Als unauffällig gelten Stanine-Werte von vier bis sechs (Horowitz et al. 2000).

Das IIP gilt wegen der standardisierten Durchführung, Auswertung und Interpretation als objektiv. Vor dem Hintergrund der mit acht Items pro Skala niedrigen Itemanzahl und der Tatsache, dass die Skalen der deutschen Fassung bereits durch die Orientierung an dem englischen Original festgelegt waren und nicht mit dem Ziel der größtmöglichen Konsistenz konstruiert wurden, bewerten Horowitz und Kollegen (2000) die interne Konsistenz mit Ausnahme der Skala LM (Cronbachs α = 0,36) als zufriedenstellend. Die Werte für Cronbachs α der ipsatierten Werte liegen zwischen α = 0,47 (Skala JK) und α = 0,64 (Skala PA) (Horowitz et al. 2000). Die Retest-Reliabilität wurde mit Hilfe einer Teilstichprobe (n = 33) der Studentenstichprobe, die nach acht Wochen das IIP wiederholte, ermittelt. Die Retest-Reliabilität beträgt r_{tt} = 0,81 (Skala BC) bis r_{tt} = 0,90 (Skala JK) (Horowitz et al. 2000).

Validierungsstudien ergaben, dass der Test diskriminative Validität aufweist, also Diagnosegruppen im klinischen Bereich unterscheiden kann (Horowitz et al. 2000; Wolf et al. 2005). Außerdem liegt faktorielle Validität vor, da die circumplexe Struktur in Faktorenanalysen mit deutschen Stichproben bestätigt werden konnte (Horowitz et al. 2000). Darüber hinaus wurde die prognostische Bedeutung interpersonaler Probleme in der Psychotherapie im Rahmen von Studien zur Kriteriumsvalidität untersucht (Horowitz et al. 2000). Im Manual zum Test werden zusätzlich die Zusammenhänge zwischen interpersonalen Problemen und Persönlichkeitsmerkmalen sowie zwischen dem IIP und weiteren klinischen Ska-

len dargestellt. Die Untersuchung dieser Zusammenhänge diente der Validierung des Konstruktes interpersonaler Probleme (Horowitz et al. 2000). Beispielsweise korrelieren die nicht ipsativen Skalenwerte des IIP mit dem GSI und den einzelnen Skalen der SCL-90-R. Bei Betrachtung der ipsativen Werte des IIP jedoch, bei denen die allgemeine Klagsamkeit, die das IIP und die SCL-90-R miterfassen, herausgerechnet wurde, zeigte sich, dass „mit dem IIP andere Probleme tangiert werden als mit der SCL-90-R“ (Horowitz et al. 2000: 36).

Das IIP ist veränderungssensitiv (Huber et al. 2007), jedoch verändern sich die Werte in geringerem Maße als bei anderen Testverfahren, zum Beispiel der SCL-90-R. Oft werden größere Effekte erst nach längeren Katamnesezeiträumen deutlich (Strauß 2003).

2.3.3 Der Beeinträchtigungs-Schwere-Score BSS

Der Beeinträchtigungs-Schwere-Score BSS dient der Ermittlung der Schwere einer psychogenen Erkrankung bzw. der Beschreibung des Grads und der Lokalisation der Beeinträchtigung durch psychogene Krankheitssymptome durch einen Experten (Schepank 1995). Schepank (1995) zählt im Manual zum BSS auf, welche Symptome er unter dem Begriff der psychogenen Symptomatik versteht. Dazu gehören unter anderem die „psychoneurotischen Symptome, wie Ängste, Depressionen, Zwangssymptome, Leeregefühl“ (Schepank 1995: 7), Persönlichkeitsstörungen, sexuelle Verhaltensabweichungen, Suchtverhalten, funktionelle psychosomatische Beschwerden, Reaktionen auf schwere Belastungen und Anpassungsstörungen, Störungen des Sozialverhaltens und Suizidversuche (Schepank 1995). Darüber hinaus fordert Schepank (1995) für die Verwendung des Begriffs „psychogen“ die Annahme einer psychischen und den Ausschluss einer primär somatischen Pathogenese.

Als Voraussetzung für eine valide Beurteilung mithilfe des BSS gilt eine vorausgegangene ausführliche Diagnostik durch einen erfahrenen Experten und der „Ausschluss einer primär somatischen Erkrankung als Ursache für die festgestellte [...] zu gewichtende Symptomatik“ (Schepank 1995: 10). Zu einer solchen umfassenden Diagnostik werden im Manual eine Anamnese, die tiefenpsychologische Aspekte und die Biographie des Patienten berücksichtigt, die ausführliche Bestimmung der aktuellen Symptome, das Erkennen der symptomauslösenden Situation und die Untersuchung der Persönlichkeitsstruktur gezählt (Schepank 1995). Der BSS kann bei Personen ab dem Vorschulalter eingesetzt werden (Schepank 1995). Er ist unabhängig von der psychologischen Theorie verwendbar und kann sowohl in der stationären und ambulanten Psychotherapie als auch in der Therapieeffekt-, Katamnese- und Grundlagenforschung genutzt werden (Schepank 1995).

Die Beeinträchtigung einer Person durch eine psychogene Erkrankung wird vom Untersucher auf den drei Dimensionen körperlich (= kö), psychisch (= psy) und sozialkommunikativ (= soko) gewichtet. Zur psychogenen körperlichen Symptomatik gehören beispielsweise Schmerzen, Schlafstörungen, innere Unruhe und Funktionsstörungen zahlreicher Organ-

systeme. Als Beispiele für die Symptomatik der psychischen Dimension gelten Depressionen, Zwangsgedanken und -handlungen, Ängste, Körperschemastörungen, Konzentrations-, Lern- und Gedächtnisstörungen. Auf der sozialkommunikativen Ebene werden Persönlichkeitsbereiche wie beispielsweise das Leistungsvermögen, das Sexual- und Suchtverhalten, die soziale Anpassung, das Kontaktverhalten und das Freizeitverhalten beurteilt (Schepank 1995). Jede dieser drei Dimensionen (kö, psy, soko) schließt fünf Stufen entsprechend den Schweregraden der Beeinträchtigung 0 = gar nicht, 1 = geringfügig, 2 = deutlich, 3 = stark und 4 = extrem ein (Schepank 1995). Die Punktewerte der Subskalen „kö", „psy" und „soko" ergeben in dieser Reihenfolge das sogenannte Profil und summiert den Gesamt-BSS (Schepank 1995). Summenwerte von 0 bis 1 gelten als „optimale Gesundheit" (Schepank 1995: 10), von 2 bis 3 als „leichte Störung" (Schepank 1995: 10), von 4 bis 5 als „deutliche Störung [mit] Krankheitswert" (Schepank 1995: 10), von 6 bis 7 als „ausgeprägte und schon ziemlich schwer beeinträchtigende Erkrankung" (Schepank 1995: 10). 8 bis 9 Punkte repräsentieren eine „außerordentlich schwere Erkrankung" (Schepank 1995: 10) und 10 bis 12 Punkte bilden „Extremgrade[ ] psychogener Erkrankungen" (Schepank 1995: 10) ab. Im Mannheimer Kohortenprojekt, einer epidemiologischen Studie an 600 Erwachsenen einer Zufallsstichprobe, wurden BSS-Summenwerte von ≥ 5 als Fall definiert (Schepank 1995). Dieser Cut-off-Wert ist für die Bestimmung der klinischen Signifikanz notwendig. Das Handbuch zum BSS enthält jedoch keine Informationen bezüglich einer Differenz zwischen den Werten zu Beginn und am Ende einer Therapie, die als statistisch signifikant zu deuten wäre. Aus diesem Grund definieren Spitzer und Kollegen (2012), basierend auf dem berechneten Reliable Change Index (RC), eine Veränderung von ≥ 2 als statistisch signifikant.

Um die Ergebnisse des BSS vergleichen zu können, muss der Beurteilungszeitraum von dem Experten angegeben werden. Die durchschnittliche Beeinträchtigung in den letzten sieben Tagen definiert die häufig verwendete Punktprävalenz. Auch eine retrospektiv ermittelte Periodenprävalenz kann mit Hilfe der durchschnittlichen BSS-Werte eines definierten Zeitintervalls bestimmt werden. Der BSS kann zur Verlaufsbeurteilung und für eine vergleichende Bestimmung, ob eine Symptomatik akut oder chronisch auftritt, genutzt werden (Schepank 1995).

Die Beurteilung durch einen unabhängigen Experten, wie sie beim BSS vorliegt, gilt als objektiv (Schepank 2003). Die Reliabilität ist abhängig von den Fähigkeiten der Experten, also den Ratern. Zur Beschreibung der Interrater-Reliabilität wird die Intraclass-Korrelation verwendet. Nach Shrout und Fleiss (1979) werden sechs Formen der Intraclass-Korrelation unterschieden. Bei der Angabe der Intraclass-Korrelationskoeffizienten ICC werden in Klammern zwei Angaben gemacht. Die erste Zahl beschreibt das Modell und informiert darüber, ob dieselbe Gruppe an Ratern alle Aspekte beurteilt und ob die Rater zufällig ausgewählt wurden (Koo und Li 2016). Es werden (1) das „One-Way Random-Effects Model" (Koo und Li 2016: 156), bei dem jeder Aspekt von unterschiedlichen, aus einer größeren Gruppe von Beurteilern zufällig ausgewählten Ratern beurteilt wird, (2)

das „Two-Way Random-Effects Model“ (Koo und Li 2016: 157), bei dem die Beurteiler zufällig aus einer Gruppe von Beurteilern mit ähnlichen Eigenschaften ausgewählt werden und dann jeder Rater jeden Aspekt beurteilt, sodass die Reliabilitätsergebnisse für alle Rater mit denselben Eigenschaften verallgemeinert werden können, und (3) das „Two-Way Mixed-Effects Model“ (Koo und Li 2016: 157), bei dem nur die ausgewählten Beurteiler von Interesse sind und die Ergebnisse nicht generalisierbar sind, unterschieden (Koo und Li 2016). Die zweite Zahl zeigt an, ob (1) die Messung eines einzelnen Raters betrachtet wird oder (k) der Mittelwert mehrerer Beurteiler (Koo und Li 2016; Shrout und Fleiss 1979). Im Manual zum BSS werden die Intraclass-Korrelationen angegeben, die eine Studie (n = 6) an Ärzten und Psychologen, die einige Patienten nach Betrachten einer Beispiel-Videosequenz und einer zugehörigen Fallvignette beurteilten, ergab. Die Intraclass-Korrelationen, als Maß für die Übereinstimmungsreliabilität eines Messinstruments, lagen für den BSS-Gesamtwert für einen mittleren einzelnen Rater einer Gruppe bei ICC (2, 1) = 0,9 und – beruhend auf den Mittelwerten einer Gruppe – bei ICC (2, n) = 0,99 (Schellberg 1995). Dies entspricht einer guten bis exzellenten Reliabilität (Koo und Li 2016).

Die Korrelationen mit dem Goldberg-Cooper-Interview, dem Freiburger Persönlichkeitsinventar, dem Gießen-Tests und der Beschwerdelisten von v. Zerssen stützen die Validität des BSS (Schepank 2003). Der BSS gilt darüber hinaus als spezifisch und sensibel für Veränderungen und als sensibel für Veränderungen der Dimension der Symptomatik (Schepank 1995).

2.3.4 Basisdokumentation und Fragen zur direkten Veränderungsmessung

Die soziodemographischen und ein Teil der klinischen Angaben der Patienten in der Studie wurden im Zuge der Basisdokumentation erfasst. Die Basisdokumentation findet im Fachklinikum Tiefenbrunn im Rahmen der Qualitätssicherung statt. Dort wird seit 2001 für das „Stuttgart-Heidelberg-Modell der aktiven Qualitätssicherung“ das Computerprogramm AKQUASI bzw. die internetbasierte Weiterentwicklung, das Fragebogen- und Monitoringprogramm Web-AKQUASI (Zimmer und Moessner 2012) genutzt. Mit dessen Hilfe können u. a. Angaben zu soziodemographischen und klinischen Merkmalen, der Motivation, den Erwartungen, der therapeutischen Beziehung, der Patientenzufriedenheit sowie den Ergebnissen der Therapie sowohl aus der Perspektive der Patienten als auch aus Sicht der Therapeuten, der Pflege und der Sozialarbeiter gesammelt, dokumentiert und rückgemeldet werden (Kordy et al. 2003; Zimmer und Moessner 2012). Die Ergebnisrückmeldung erfolgt in AKQUASI entweder als einzeitige Messdarstellung oder als Verlaufsdarstellung oder als standardisierte Ergebnisbewertung, welche auf veröffentlichten Referenzwerten beruht (Kordy et al. 2003).

Für die direkte Veränderungsmessung schätzten sowohl der Bezugstherapeut als auch der Patient zum Zeitpunkt der Entlassung auf einer fünfstufigen Ratingskala ein, wie sich die

Symptomatik des Patienten im Verlauf der Therapie verändert hat. Dabei haben sie die Wahl zwischen folgenden Stufen: 1 = „deutlich verschlechtert“, 2 = „leicht verschlechtert“, 3 = „weder noch“, 4 = „leicht gebessert“ oder 5 = „deutlich gebessert“.

2.4 Datenverarbeitung und statistische Verfahren

Zur statistischen Auswertung wurden die Computerprogramme IBM SPSS Statistics 23 und Statistika 12 verwendet. Im Rahmen der indirekten Veränderungsmessung wurden der Global Severity Index (GSI) der SCL-90-R, der 64-Item-Gesamtwert (IIP_{ges}) des IIP und der BSS-Gesamtwert zur Analyse genutzt. Es wurden für alle Tests nicht-transformierte Werte angegeben. Die Einschätzung des Behandlungserfolgs anhand der Veränderung der Symptomatik im Sinne der direkten Verlaufsmessung wurde zusätzlich untersucht. Es wurde für diese Skalen Intervallskalenniveau angenommen. Im Folgenden werden die statistischen Verfahren aufgezeigt, die zur Prüfung der Repräsentativität der Stichprobe, zur Beantwortung der Forschungsfragen und zur Untersuchung, ob soziodemographische Unterschiede zwischen den untersuchten Gruppen vorliegen und welchen Einfluss diese auf Misserfolge haben, verwendet wurden. Signifikante, hoch signifikante und höchst signifikante Ergebnisse werden wie folgt gekennzeichnet: $p < 0{,}05$ *, $p < 0{,}01$ **, $p < 0{,}001$ ***.

2.4.1 Methoden zur Stichprobenbeschreibung und Prüfung der Repräsentativität

Zur Charakterisierung der Stichprobe wurden die Häufigkeiten einiger soziodemographischer und klinischer Merkmale beschrieben. Um zu prüfen, ob sich die Teil- und die Gesamtstichprobe unterscheiden, wurden soziodemographische Parameter und die im Rahmen der Forschungsfragen untersuchte Variable „Vorliegen einer Persönlichkeitsstörung“ und deren Differenzierung in Cluster-B- und Cluster-C-Persönlichkeitsstörung der ausgewählten Stichprobe mit den Angaben der insgesamt im Zeitraum vom 01.01.2005 bis zum 31.12.2014 in Tiefenbrunn behandelten Patienten verglichen. Der Vergleich erfolgte für das Alter mittels t-Test und für die nominal- und ordinalskalierten Variablen mit χ^2-Tests. Mithilfe von t-Tests können Mittelwertsunterschiede auf Signifikanz geprüft werden (Kuckartz et al. 2013). Falls bei der Anwendung von t-Tests in dieser Arbeit der Levene-Test darauf hinwies, dass keine Varianzhomogenität gegeben war, wurde der diesbezüglich robustere Welch-Test angegeben. χ^2-Tests können Häufigkeitsunterschiede auf Signifikanz prüfen (Bühner und Ziegler 2009) und somit testen, ob in der Grundgesamtheit ein Zusammenhang zwischen zwei Merkmalen besteht (Duller 2013). Mithilfe von χ^2 kann die Abweichung zwischen beobachteten und erwarteten Häufigkeiten in einer Maßzahl beschrieben werden (Kuckartz et al. 2013).

2.4.2 Methoden zur Untersuchung der ersten Forschungsfrage

Zur Beantwortung der Forschungsfrage 1 wurden die Anteile an Patienten bzw. Therapeuten bestimmt, die die Symptomatik des Patienten als deutlich verschlechtert, leicht ver-

schlechtert, weder gebessert noch verschlechtert, leicht gebessert oder deutlich gebessert einschätzten (direkte Veränderungsmessung). Außerdem wurden die Häufigkeitsverteilungen der Ergebnisgruppen Heilung, Besserung, Nonresponse, Verschlechterung, klinisch bedeutsame Verschlechterung und Pseudogesund (s. 1.3.2) für die Ergebnisse der SCL-90-R, des IIP und des BSS ermittelt.

Zur Bildung der Ergebnisgruppen (s. 1.3.2) wurden in dieser Arbeit für die SCL-90-R die von Franke (2002) vorgeschlagene kritische Differenz von ± 0,3 (Bestimmung der statistischen Signifikanz) und der Cut-off-Wert von c = 0,7, der den funktionalen vom dysfunktionalen Bereich trennt (Bestimmung der klinische Signifikanz), verwendet. Für das IIP wurden eine kritische Differenz von ± 0,4243 und ein Cut-off von c = 1,51, die gemäß den Formeln in Kapitel 1.3.2 berechnet wurden, genutzt. Für die Gruppeneinteilung mit Hilfe des BSS wurden der von Schepank (1995) definierte Cut-off-Wert von c = 5 und die von Spitzer und Kollegen (2012) berechnete kritische Differenz von ± 2 angewendet. Diese Annahmen liegen den Ergebnissen aller drei Forschungsfragen zugrunde.

2.4.3 Methoden zur Untersuchung der zweiten und dritten Forschungsfrage

Zur Beantwortung der Forschungsfragen Nr. 2 und 3 wurden in der direkten Veränderungsmessung zur Untersuchung von dichotomen Faktoren t-Tests für unabhängige Stichproben durchgeführt. Die t-Tests wurden für die Angaben der Therapeuten und der Patienten eingesetzt.

Für die indirekte Messung wurden jeweils die Ergebnisse aller drei Tests (SCL-90-R, BSS, IIP) untersucht. Die Ergebnisgruppen wurden für die Untersuchungen sowohl dichotom (Erfolg vs. Misserfolg) als auch dreistufig (Erfolg vs. Nonresponse vs. Verschlechterung) verwendet. Dazu wurden die Ergebnisgruppen Heilung und Besserung zur Gruppe Erfolg zusammengefasst. Die Gruppen Verschlechterung und klinisch bedeutsame Verschlechterung bildeten die Gruppe Verschlechterung. Die Gruppe Nonresponse wurde entweder eigenständig betrachtet oder zusammen mit der Gruppe der Verschlechterten zur Misserfolgsgruppe gezählt. Ob sich Männer und Frauen oder Personen mit und ohne Persönlichkeitsstörungen oder Personen mit Persönlichkeitsstörungen der Cluster B und C in der indirekten Veränderungsmessung hinsichtlich Erfolg und Misserfolg bzw. Erfolg, Nonresponse und Verschlechterung unterscheiden, wurde aufgrund des Nominalskalenniveaus mithilfe von χ^2-Tests untersucht. Bei den Ergebnissen wurden teilweise die standardisierten Residuen angegeben. Diese berechnen sich aus der Differenz der bei statistischer Unabhängigkeit zu erwartenden Häufigkeit und der tatsächlich gemessenen Häufigkeit dividiert durch die Wurzel der erwarteten Häufigkeit (Bühl 2006). Da sich der χ^2-Wert aus der Summe der quadrierten standardisierten Residuen berechnet, geben die standardisierten Residuen darüber Auskunft, wie hoch der Beitrag eines Feldes einer Kreuztabelle zum χ^2-Wert und zu einem ggf. signifikanten Ergebnis ist (Bühl 2016). Standardisierte Residuen, die größer oder gleich zwei sind, gelten als signifikant (Bühl 2016). Darüber hinaus kann

beispielsweise das Zusammenhangsmaß Cramers V zur Angabe der Stärke des Zusammenhangs mithilfe von Zahlenwerten im Bereich zwischen 0 und 1 angegeben werden (Bühl 2016; Kuckartz et al. 2013). Bei der Durchführung der χ^2-Tests wurde stets darauf geachtet, dass die Voraussetzung erfüllt war, dass erwartete Häufigkeiten kleiner als 5 in maximal einem Fünftel der Zellen einer Tabelle vorlagen (Bühl 2016; Kuckartz et al. 2013).

Die Summenscores bei Entlassung und bei Aufnahme wurden auf Mittelwertsunterschiede zwischen den Gruppen der unterschiedlichen Geschlechter, zwischen den Patienten mit Persönlichkeitsstörungen vs. denen ohne und zwischen den Clustern B und C der Persönlichkeitsstörungen geprüft. Dies geschah mithilfe zweifaktorieller Varianzanalysen mit Messwiederholung auf einem Faktor. Hierbei wurden die Haupteffekte Messzeitpunkt und Gruppenzugehörigkeit sowie der Wechselwirkungseffekt zwischen diesen beiden Haupteffekten untersucht. Dazu wurden die Effektstärken $\eta_{partial}^2$ angegeben. $\eta_{partial}^2$ beschreibt den Anteil des systematischen Effekts an der Variation, die durch den Effekt selber und die unsystematische Variation zustande kommt (Bühner und Ziegler 2009), also den erklärten Varianzanteil an der Varianz, die nicht durch andere untersuchte Faktoren erklärt werden kann (Döring und Bortz 2016c). $\eta_{partial}^2$ lässt sich allgemein nach folgender Formel berechnen: $\eta_{\text{partial}}^2 = \frac{QS_{\text{Effekt}}}{QS_{\text{Effekt}} + QS_{\text{Fehler}}}$ mit QS als Quadratsumme des Effekts bzw. des Fehlers. Genauer betrachtet gibt die Effektstärke der Gruppenzugehörigkeit den Anteil an Unterschieden zwischen den Personen an, der durch die Gruppenzugehörigkeit verursacht wird und berechnet sich wie folgt: $\eta_{\text{partial}}^2 = \frac{QS_{\text{Gruppenfaktor}}}{QS_{\text{zw Vpn}}} = \frac{QS_{\text{Gruppenfaktor}}}{QS_{\text{Gruppenfaktor}} + QS_{\text{inn S}}}$ mit $QS_{\text{Gruppenfaktor}}$ (Quadratsumme des Gruppenfaktors) als systematischer Gruppenfaktoreffekt und $QS_{\text{zw Vpn}}$ als Quadratsumme zwischen den Versuchspersonen, die sich aus der Quadratsumme des Gruppenfaktors und $QS_{\text{inn S}}$ (Quadratsumme innerhalb der Stichproben) als unsystematischer Variation innerhalb der Stichproben zusammensetzt (Bühner und Ziegler 2009). Für die Effektstärken der Messwiederholung und des Wechselwirkungseffekts gilt äquivalent: $\eta_{\text{partial}}^2 = \frac{QS_{\text{Effekt}}}{QS_{\text{Effekt}} + QS_{\text{Residuum}}}$, mit QS_{Effekt} als Quadratsumme der Messwiederholung oder der Wechselwirkung und QS_{Res} als Quadratsumme des Residuums. Letztgenannte kommt aus Messfehlern und Wechselwirkungen zwischen dem Messzeitpunkt und spezifischen Eigenschaften der einzelnen Personen zustande (Bühner und Ziegler 2009). Um die Ergebnisse der interessierenden Gruppen untereinander und mit den Ergebnissen anderer Studien, auch wenn diese ein anderes Studiendesign aufweisen oder andere abhängige Maße verwenden, zu vergleichen, wurden ergänzend Effektstärken d ermittelt (Bühner und Ziegler 2009). Die Effektstärken d der Stichproben werden nach der Formel $d = (M_{\text{Aufnahme}} - M_{\text{Entlassung}}) / SD_{\text{Aufnahme}}$ berechnet (Spitzer et al. 2015), wobei M den Mittelwert und SD die Standardabweichung darstellt.

2.4.4 Methoden zur Untersuchung soziodemographischer Aspekte

Mit dem Ziel, einschätzen zu können, ob die Unterschiede zwischen den untersuchten Gruppen und den Therapieergebnissen auf die untersuchte Variable des Geschlechts oder das Vorliegen einer Persönlichkeitsstörung zurückzuführen sind, oder eventuell andere Variablen wie soziodemographische Unterschiede als Ursache in Betracht kommen, wurde untersucht, ob sich die in der zweiten und dritten Forschungsfrage analysierten Gruppen hinsichtlich soziodemographischer Aspekte unterscheiden.

2.4.4.1 Prüfung auf das Vorliegen von soziodemographischen Unterschieden zwischen den analysierten Gruppen

Die Untersuchung auf das Vorliegen von Unterschieden hinsichtlich soziodemographischer Aspekte der analysierten Gruppen erfolgte für kategoriale Variablen auf Nominalskalenniveau mithilfe von χ^2-Tests und für kontinuierliche Merkmale mittels t-Tests. Die kategorialen soziodemographischen Variablen wurden dafür wie in Tabelle 7 dargestellt zusammengefasst und differenziert.

Tabelle 7: Soziodemographische Merkmale mit ihren Ausprägungen

Soziodemographisches Merkmal	Merkmalsausprägung
Geschlecht	männlich
	weiblich
Familienstand	ledig
	verheiratet
	geschieden / verwitwet / getrennt lebend
höchster Schulabschluss	Hauptschulabschluss
	Realschulabschluss / Mittlere Reife
	Abitur / Fachhochschulreife
	Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige
höchster Berufsabschluss	noch in der Berufsausbildung
	Lehre
	Meister / Fachschule
	Fachhochschule / Universität
	ohne Abschluss
	sonstiger Abschluss
Berufstätigkeit vor Aufnahme	Arbeiter
	Facharbeiter
	einfacher / mittlerer Angestellter / Beamter
	höherer Angestellter / Beamter
	selbständig
	Schüler / Student / Auszubildender / Umschüler
	Rentner / Pensionär
	ohne Beruf / Hausfrau / -mann
Beschäftigung vor Aufnahme	voll erwerbstätig
	teilzeitbeschäftigt
	nicht erwerbstätig
	arbeitslos
Berentung	ja
	nein
Arbeitsfähigkeit bei Aufnahme	arbeits- / schulfähig / teilweise arbeits- / teilweise schulfähig
	arbeits- / schul- / berufs- / erwerbsunfähig

2.4.4.2 Prüfung des Einflusses soziodemographischer Parameter auf das Therapieergebnis

Um der Frage nachzugehen, welchen Einfluss die o. g. soziodemographischen Parameter, hinsichtlich derer sich die untersuchten Gruppen teilweise signifikant unterscheiden (s. 3.5), auf die Auftretenswahrscheinlichkeit von Erfolg oder Misserfolg haben, wurde eine logistische Regression durchgeführt.

Im Folgenden soll zunächst das Modell der logistischen Regression vorgestellt werden. Die logistische Regression stellt eine geeignete Methode zur Analyse des Einflusses kategorialer

– nominal- oder ordinalskalierter – und metrischer unabhängiger Variablen auf eine nominalskalierte abhängige Variable mit zwei Ausprägungen (in diesem Fall Erfolg (y = 0) vs. Misserfolg (y = 1), gemessen mithilfe der SCL-90-R) dar (Backhaus et al. 2011; Field 2013; Fromm 2005). Sie zielt darauf ab, mithilfe der logistischen Funktion p abhängig von verschiedenen Faktoren die Wahrscheinlichkeit für das Auftreten des Ereignisses y = 1, hier Misserfolg, vorherzusagen (Backhaus et al. 2011). Da die logistische Funktion p eine Wahrscheinlichkeitsbeziehung zwischen den unabhängigen Variablen und dem Ereignis y = 1 aufbaut, wird sie auch „linking function" genannt (Backhaus et al. 2011). In Abbildung 3 sind die Zusammenhänge zwischen den Parametern der logistischen Regression dargestellt. Es wird ein nicht-linearer Zusammenhang zwischen den unabhängigen Variablen (x_{jk}) sowie der Eintrittswahrscheinlichkeit der dichotomen abhängigen Variable (P(y = 1)) und ein linearer Zusammenhang zwischen den unabhängigen Variablen und der aggregierten Einflussstärke z_k, die im Exponenten der linking-Funktion verwendet wird, angenommen (Backhaus et al. 2011).

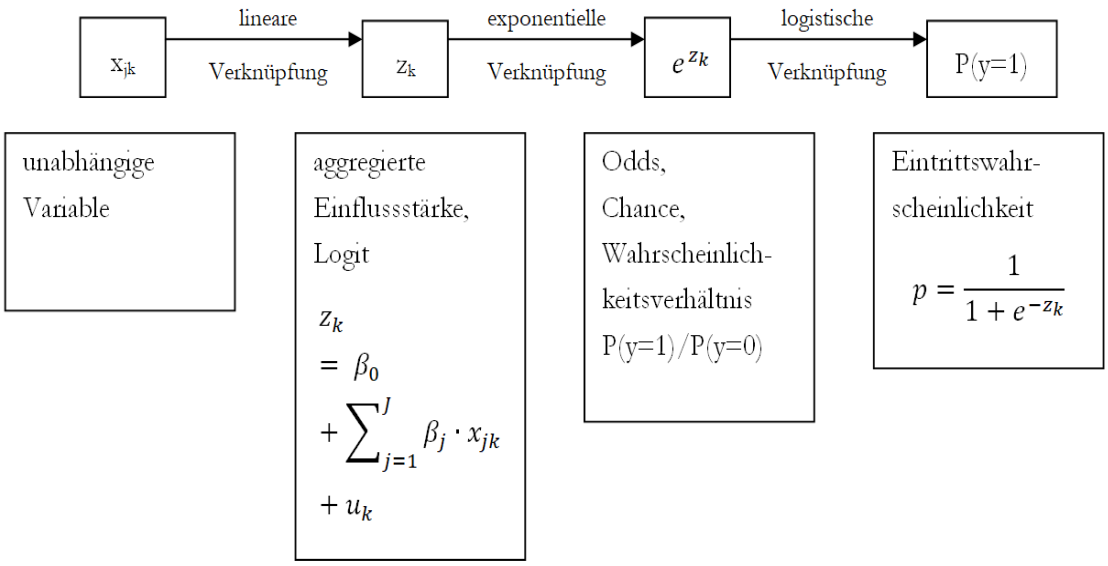

Abbildung 3: Zusammenhänge zwischen den Parametern der logistischen Regression, modifiziert nach Backhaus et al. (2011)

Bei der Berechnung der aggregierten Einflussstärke z_k, des sogenannten Logits, verdeutlichen der Parameter β_0 und die Regressionskoeffizienten β_j (= Logit-Koeffizienten) die Einflussstärke der unabhängigen Variablen x_{jk} auf die Höhe der Wahrscheinlichkeit für das Eintreten des Ereignisses y = 1. Anders als bei der linearen Regression sind unmittelbar jedoch nur die Richtungen des Einflusses der unabhängigen Variablen auf die Eintrittswahrscheinlichkeit erkennbar. Positive Koeffizienten erhöhen die Auftretenswahrscheinlichkeit für y = 1, negative Koeffizienten verringern die Wahrscheinlichkeit für dieses Ereignis bei steigendem x_{jk} (Backhaus et al. 2011). Die Erhöhung einer unabhängigen Variablen (x_{jk}) um eine Einheit führt linear zur Veränderung des logarithmierten Wahrscheinlich-

keitsverhältnisses (Logit) (Backhaus et al. 2011). Die Formel zur Berechnung der Wahrscheinlichkeit p für das Eintreten des Ereignisses y = 1 (Misserfolg) für eine Person k ergibt sich aus $p(y=1)=\frac{e^{z_k}}{1+e^{z_k}}$ bzw. $p=\frac{1}{1+e^{-z_k}}$ mit $z_k = \beta_0 + \sum_{j=1}^{J} \beta_j \cdot x_{jk} + u_k$ (Backhaus et al. 2011) und kann dementsprechend nach der Formel $p_k = \frac{1}{1+e^{-(\beta_0+\sum_{j=1}^{J}\beta_j \cdot x_{jk}+u_k)}}$ berechnet werden. Dabei ist j die Anzahl der unabhängigen Variablen und u_k ein Fehlerterm.

Mithilfe von Effektkoeffizienten können Aussagen über die Höhe der Einflussstärken getroffen werden (Backhaus et al. 2011). Der Effektkoeffizient (e^b) ist der „Faktor, um den sich die Chance für das Eintreten von y = 1 ändert, wenn sich die unabhängige Variable um eine Einheit ändert (bei metrischen Variablen) bzw. wenn man die betrachtete Ausprägung einer kategorialen Variable mit ihrer Referenzkategorie vergleicht“ (Fromm 2005: 24). Bei Werten kleiner als 1 sinkt die Chance, bei Werten über 1 steigt die Chance und bei einem Wert von genau 1 bleibt die Chance gleich (Fromm 2005). Der Effektkoeffizient entspricht der Odds-Ratio (Backhaus et al. 2011; Fromm 2005), also dem Quotienten aus den Odds nach der Veränderung um eine Einheit und vor der Veränderung (Field 2013).

Nach dieser allgemeinen Vorstellung des Modells der logistischen Regression wird an dieser Stelle speziell auf das Vorgehen in dieser Arbeit eingegangen. Zum Ausschluss von Multikollinearität, also dass ein Prädiktor aus anderen Prädiktoren vorhersagbar ist, was das Ergebnis der Regressionsanalyse verzerren würde, wurden die Korrelationen zwischen den unabhängigen Variablen untersucht (Kuckartz et al. 2013).

Die kategorialen Variablen mit mehr als zwei (k > 2) Ausprägungen wurden in k - 1 verschiedene Dummy-Variablen umkodiert (Fromm 2005; Kuckartz et al. 2013). Dummy-Variablen sind binäre Variablen, die für die einzelnen Ausprägungen von nominalen oder ordinalen Variablen erstellt werden und nur die Ausprägungen 0 = „nein“ und 1 = „ja“ aufweisen (Kuckartz et al. 2013). Als Referenzkategorie gilt die Ausprägung der ursprünglichen Variable, für die keine Dummy-Variable gebildet wird. Sie wird auf inhaltlichen Überlegungen basierend festgelegt, beispielsweise so, dass ein Extremgruppenvergleich möglich ist (Fromm 2005). Die Dummy-Variablen und die Referenzkategorien sind in Tabelle 30 dargestellt.

Die Aufnahme der unabhängigen Variablen in das Modell erfolgte vorwärts schrittweise. Das heißt, in jedem Schritt wird die Variable mit dem in diesem Schritt berechneten höchsten und gleichzeitig signifikanten Chi-Quadrat-Wert integriert. Eine solche Aufnahme an Variablen erfolgt bis alle unabhängigen Variablen in das Modell aufgenommen worden sind oder keine signifikanten Chi-Quadrat-Werte der übrigen unabhängigen Variablen mehr vorliegen und das Modell nicht mehr verbessert werden kann. Die zu einer unabhängigen Variable gehörenden Dummy-Variablen werden als „Variablenset“ innerhalb eines Schrittes in das Modell eingeordnet (Fromm 2005).

Die logistische Regressionsanalyse zielt darauf, Gruppenzugehörigkeiten zu prognostizieren oder Gruppenunterschiede zu erklären (Fromm 2005). „Ein Fall wird der Gruppe mit der Ausprägung ‚1' zugeordnet, wenn die prognostizierte Wahrscheinlichkeit, der Gruppe 1 anzugehören, mindestens so groß wie der Cut-Wert ist" (Fromm 2005: 15-16). Die Ausprägung „1" der abhängigen Variablen kodiert in dieser Arbeit Misserfolg und die Ausprägung „0" Erfolg. Der Cut-off Wert ist gleich dem Anteil der Fälle mit der Ausprägung „1" (Fromm 2005), also dem Anteil der Patienten mit Misserfolg an allen Patienten, die in der logistischen Regression betrachtet wurden.

3 Ergebnisse

In diesem Kapitel wird zuerst die untersuchte Stichprobe charakterisiert. Im Anschluss werden die Häufigkeiten für Nonresponse und Verschlechterungen für die einzelnen Messmethoden angegeben und es wird dargestellt, ob und bei welchen Erhebungsarten Unterschiede zwischen den untersuchten Gruppen bezüglich Misserfolgen und hinsichtlich der Symptombelastung, der Stärke interpersonaler Probleme und der Beeinträchtigungsschwere vorlagen. Darüber hinaus wird beschrieben, inwiefern sich die untersuchten Gruppen hinsichtlich soziodemographischer Aspekte unterschieden.

3.1 Beschreibung der Stichprobe

Es erfüllten 3862 Patienten (59,6 %) die Eingangskriterien. Von den 6481 Patienten, die insgesamt im Zeitraum vom 01.01.2005 bis zum 31.12.2014 im Fachklinikum Tiefenbrunn behandelt und erfasst wurden, erfüllten neun Patienten (0,1 %) das Mindestalter von 18 Jahren nicht, 1361 Patienten der Gesamtstichprobe (21,0 %) wurden aufgrund einer Behandlungsdauer von unter 6 Wochen ausgeschlossen und für 2232 Patienten (34,4 %) lagen keine Werte für den GSI der SCL-90-R zum Aufnahme- und Entlassungszeitpunkt vor.

Für alle 3862 Patienten lagen zum Aufnahme- und zum Entlassungszeitpunkt SCL-90-R Werte vor, da deren Vollständigkeit ein Einschlusskriterium war. Für 3814 Patienten wurden zu diesen beiden Zeitpunkten auch die Werte des IIP dokumentiert. Vollständige Werte des BSS bei Aufnahme und bei Entlassung waren von 2148 Patienten vorhanden.

3.1.1 Soziodemographische Merkmale

Das Durchschnittsalter der Patienten in der Stichprobe lag bei Aufnahme bei 35,5 Jahren (SD = 11,8 Jahre, Bereich: 18 bis 72 Jahre). Weitere soziodemographische Charakteristika der Untersuchungsstichprobe, wie die Geschlechterverteilung, der Familienstand, der höchste Schul- und Berufsabschluss, die Berufstätigkeit und Beschäftigung vor der Aufnahme, Angaben zur Berentung und die Arbeitsfähigkeit sind in Tabelle 8 dargestellt.

Tabelle 8: Soziodemographische Charakteristika der Untersuchungsstichprobe (n = 3862)

Merkmal	n	%
Geschlecht		
männlich	1440	37,3
weiblich	2422	62,7
Familienstand		
ohne Angaben	1	0,03
ledig	2482	64,3
verheiratet	798	20,7
geschieden	432	11,2
verwitwet	41	1,1
getrennt lebend	108	2,8
höchster Schulabschluss		
Hauptschulabschluss	635	16,4
Realschulabschluss / Mittlere Reife	1291	33,4
Abitur / Fachhochschulreife	1682	43,6
Sonderschulabschluss	16	0,4
ohne Schulabschluss	47	1,2
noch in der Schule	48	1,2
sonstiger Abschluss	132	3,4
ohne Angabe	11	0,3
höchster Berufsabschluss		
noch in der Berufsausbildung	356	9,2
Lehre	1261	32,7
Meister / Fachschule	282	7,3
Fachhochschule / Universität	706	18,3
ohne Abschluss	780	20,2
sonstiger Abschluss	456	11,8
ohne Angabe	21	0,5
Berufstätigkeit vor Aufnahme		
Arbeiter	383	9,9
Facharbeiter	187	4,8
einfacher / mittlerer Angestellter / Beamter	879	22,8
höherer Angestellter / Beamter	265	6,9
selbständig	145	3,8
Schüler / Student	580	15,0
Auszubildender / Umschüler (nicht Student)	224	5,8
Hausfrau / -mann	174	4,5
Rentner / Pensionär	401	10,4
ohne Beruf	608	15,7
ohne Angabe	16	0,4

Tabelle 8 (Fortsetzung): Soziodemographische Charakteristika der Untersuchungsstichprobe (n = 3862)

Merkmal	n	%
Beschäftigung vor Aufnahme		
voll erwerbstätig	1248	32,3
teilzeitbeschäftigt	600	15,5
nicht erwerbstätig	1100	28,5
arbeitslos	888	23,0
ohne Angabe	26	0,7
Berentung		
ja	498	12,9
nein	3348	86,7
ohne Angabe	16	0,4
Arbeitsfähigkeit bei Aufnahme		
arbeits- / schulfähig	424	11,0
teilweise arbeits- / schulfähig	244	6,3
arbeits- / schulunfähig	1582	41,0
berufsunfähig	42	1,1
erwerbsunfähig	180	4,7
trifft nicht zu (z. B. wg. Altersrente)	70	1,8
nicht beurteilbar	74	1,9
ohne Angabe	1246	32,3

3.1.2 Klinische Charakteristika

Die Patienten befanden sich durchschnittlich 89,7 Tage (SD = 30,1 Tage, minimal 42 Tage, maximal 252 Tage) in psychotherapeutischer Behandlung im Fachklinikum Tiefenbrunn. Es wurden bis zu sieben Entlassungsdiagnosen pro Patient gestellt und dokumentiert. Im Durchschnitt wiesen die Patienten 3,03 (SD = 1,36) Diagnosen auf. Tabelle 9 stellt die Verteilung der ICD-10-Diagnosegruppen dar.

Tabelle 9: ICD-10-Diagnosen der Patienten der Untersuchungsstichprobe (n = 3862)

ICD-10-GM Code		n	%
F00-F09	Organische, einschließlich symptomatischer psychischer Störungen	5	0,1
F10-F19	Psychische und Verhaltensstörungen durch psychotrope Substanzen	484	12,5
F20-F29	Schizophrenie, schizotype und wahnhafte Störungen	235	6,1
F30-F39	Affektive Störungen	3243	84,0
F40-F48	Neurotische, Belastungs- und somatoforme Störungen	2841	73,6
F50-F59	Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren	562	14,6
F60-F69	Persönlichkeits- und Verhaltensstörungen	2034	52,7
F70-F79	Intelligenzstörung	5	0,1
F80-F89	Entwicklungsstörungen	5	0,1
F90-F98	Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend	31	0,8

Die in absteigender Reihenfolge häufigsten zehn Diagnosegruppen waren: F33 rezidivierende depressive Störungen (n = 2717), F60 spezifische Persönlichkeitsstörungen (n = 1873), F40 phobische Störungen (n = 1786), F43 Reaktionen auf schwere Belastungen und Anpassungsstörungen (n = 1573), F45 somatoforme Störungen (n = 503), F50 Essstörungen (n = 489), F32 depressive Episode (n = 445), F44 dissoziative Störungen (n = 376), F42 Zwangsstörung (n = 338), F41 andere Angststörungen (n = 329).

Tabelle 10 zeigt, wie viele Patienten der Stichprobe zuvor eine psychotherapeutische, psychiatrische oder andere Vorbehandlung erhalten hatten. Der Anteil an Patienten, für die entweder eine psychotherapeutische oder psychiatrische oder beide Formen der Vorbehandlung dokumentiert waren, betrug 63,2 % (n = 2439) von allen 3862 Patienten der Stichprobe ohne Berücksichtigung fehlender Werte. Die Problematik der fehlenden Werte wird in Kapitel 4.3.1.2 näher erläutert.

Tabelle 10: Vorbehandlung der Patienten

	Psychotherapeutische Vorbehandlung		Psychiatrische Vorbehandlung		Andere Vorbehandlung	
	dokumentierte Anzahl (n)	davon bejahter Anteil (%)	dokumentierte Anzahl (n)	davon bejahter Anteil (%)	dokumentierte Anzahl (n)	davon bejahter Anteil (%)
Vorbehandlung der Erkrankung, allgemein	2174	96,2	1664	97,8	978	98,2
Vorbehandlung innerhalb der letzten 12 Monate	2169	84,1	1696	83,5	1103	78
Ambulante Vorbehandlung der Erkrankung	1926	96,6	1210	98,3	847	99,1
Stationäre Vorbehandlung der Erkrankung	717	90	817	97,9	217	94,9

In Tabelle 11 werden die Häufigkeiten von Suizidalität, Suizidversuchen und parasuizidalem Verhalten der Patienten, beruhend auf den Einschätzungen der Therapeuten, aufgelistet. Unter Suizidalität wird „die Summe aller Denk-, Verhaltens- und Erlebensweisen von Menschen, die in Gedanken, durch aktives Handeln oder passives Unterlassen oder durch Handelnlassen den eigenen Tod anstreben bzw. als mögliches Ergebnis einer Handlung in Kauf nehmen“ (Wolfersdorf 2008: 1321), verstanden.

Tabelle 11: Therapeutenangaben zu Suizidalität, Suizidversuchen und parasuizidalem Verhalten der Patienten

	n	%
Suizidalität im Vorfeld der Aufnahme	1262	47,6
Chronische Suizidalität		
nur früher	184	6,9
bis zur Aufnahme	881	33,1
Gesicherter Suizidversuch im Vorfeld der Aufnahme	192	7,3
Suizidversuche in der Vorgeschichte		
nur früher	416	15,7
bis zur Aufnahme	145	5,5
Parasuizidales Verhalten in der Vorgeschichte		
nur früher	304	11,5
bis zur Aufnahme	267	10,1

Die Patienten, von denen mindestens ein Suizidversuch in der Vorgeschichte bekannt war und von denen die Anzahl der Suizidversuche erfasst wurde (n = 515), begingen im Mittel 2,21 Suizidversuche (SD = 2,3, Bereich: 1-23 Suizidversuche).

3.1.3 Vergleich zwischen der untersuchten Stichprobe und der Stichprobe vor Anwendung der Einschlusskriterien

Die Patienten der untersuchten Stichprobe nach Anwendung der Ein- und Ausschlusskriterien (n = 3862; im Folgenden als Teilstichprobe bezeichnet) unterschieden sich nicht signifikant von allen, die im Zeitraum vom 01.01.2005 bis zum 31.12.2014 im Fachklinikum Tiefenbrunn behandelt wurden (n = 6481; im Folgenden als Gesamtstichprobe bezeichnet), hinsichtlich der in den Forschungsfragen 2 und 3 untersuchten Variablen Geschlecht (χ^2 = 1,114, p = 0,291, n. s.) und dem Vorliegen einer Persönlichkeitsstörung (χ^2 = 2,341, p = 0,126, n. s.). In der Teilstichprobe befanden sich 37,3 % Männer (n = 1440) und 62,7 % Frauen (n = 2422), während in der Gesamtstichprobe 38,3 % Männer (n = 2484) und 61,7 % Frauen (n = 3997) waren. In der Teilstichprobe wiesen 48,0 % (1853 Patienten) und in der Gesamtstichprobe 46,4 % der Patienten (3009 Patienten) eine Persönlichkeitsstörung auf. Der Anteil an Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörung unterschied sich in beiden Stichproben jedoch signifikant voneinander (χ^2 = 13,408, p = 0,001**). 23,3 % der Patienten der Teilstichprobe (898 Patienten) und 24,8 % der Patienten der Gesamtstichprobe (1607 Patienten) besaßen eine Cluster-B-Persönlichkeitsstörung. 22,8 % der Patienten der Teilstichprobe (882 Patienten) und 19,9 % der Patienten der Gesamtstichprobe (1288 Patienten) wiesen eine Cluster-C-Persönlichkeitsstörung auf. Für 53,9 % der Patienten der Teilstichprobe (2082 Patienten) und 55,3 % der Patienten der Gesamtstichprobe (3586 Patienten) wurde keine Cluster-B- oder Cluster-C-Persönlichkeitsstörung dokumentiert. Das standardisierte Residuum für eine Cluster-C-Persönlichkeitsstörung in der Teilstichprobe betrug 2,5 und zeigt, dass Patienten mit Cluster-C-Persönlichkeitsstörung signifikant häufiger in der Teilstichprobe waren als erwartet. Das standardisierte Residuum für eine Cluster-C-Persönlichkeitsstörung in der Gesamtstichprobe maß -1,9. Die standardisierten Residuen für eine Cluster-B-Persönlichkeitsstörung wiesen mit -1,2 für die Teilstichprobe und 0,9 für die Gesamtstichprobe niedrigere, nicht signifikante Beträge auf. Für das Vorliegen keiner Cluster-B- oder Cluster-C-Persönlichkeitsstörungen waren die standardisierten Residuen noch kleiner: -0,7 für die Teilstichprobe und 0,6 für die Gesamtstichprobe.

Hinsichtlich soziodemographischer Aspekte unterschieden sich die Teil- und die Gesamtstichprobe nur im höchsten Schulabschluss signifikant voneinander. Der Anteil an Patienten mit Abitur war in der Teilstichprobe signifikant höher. Der Anteil an Patienten, die noch zur Schule gingen oder die einen Haupt- oder Sonderschulabschluss oder keinen Abschluss besaßen, war hingegen in der Gesamtstichprobe signifikant höher (s. Tabelle 12). Die standardisierten Residuen betrugen für die Zugehörigkeit zur Teilstichprobe für das

Abitur bzw. die Fachhochschulreife 1,6 und für den Hauptschulabschluss -1,3. Für die Zugehörigkeit zur Gesamtstichprobe lag das standardisierte Residuum für das Abitur bzw. die Fachhochschulreife bei -1,3.

Die Patienten der Teilstichprobe waren bei Aufnahme durchschnittlich 35,47 Jahre alt (SD = 11,76 Jahre) und die Patienten der Gesamtstichprobe durchschnittlich 35,44 Jahre (SD = 12,24 Jahre). Dieser Unterschied war nicht signifikant (t = 0,13, p = 0,901, n. s.). Die Ergebnisse für weitere soziodemographische Parameter sind in der Tabelle 12 dargestellt.

Tabelle 12: Vergleich soziodemographischer Parameter der Teilstichprobe mit den Parametern der Gesamtstichprobe

Merkmal	Stichprobenzugehörigkeit				Statistik	
	Teilstichprobe		Gesamtstichprobe		χ^2	p
	n	%	n	%		
Geschlecht					1,114	0,291
männlich	1440	37,3	2484	38,3		
weiblich	2422	62,7	3997	61,7		
Familienstand					1,838	0,399
ledig	2482	64,3	4119	63,6		
verheiratet	798	20,7	1321	20,4		
geschieden / verwitwet / getrennt	581	15,0	1040	16,0		
höchster Schulabschluss					9,504	0,023
Hauptschulabschluss	635	16,5	1059	18,0		
Realschulabschluss / Mittlere Reife	1291	33,5	2010	34,1		
Abitur / Fachhochschulreife	1682	43,7	2408	40,9		
Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige	243	6,3	416	7,1		
höchster Berufsabschluss					8,216	0,145
noch in Berufsausbildung	356	9,3	554	9,4		
Lehre	1261	32,8	1929	32,8		
Meister / Fachschule	282	7,3	413	7,0		
Fachhochschule / Universität	706	18,4	989	16,8		
ohne Abschluss	780	20,3	1313	22,3		
sonstiger Abschluss	456	11,9	681	11,6		

Tabelle 12 (Fortsetzung): Vergleich soziodemographischer Parameter der Teilstichprobe mit den Parametern der Gesamtstichprobe

Merkmal	Stichprobenzugehörigkeit				Statistik	
	Teilstichprobe		Gesamtstichprobe		χ^2	p
	n	%	n	%		
Berufstätigkeit vor Aufnahme					9,87	0,196
Arbeiter	383	10,0	568	9,7		
Facharbeiter	187	4,9	288	4,9		
einfacher / mittlerer Angestellter / Beamter	879	22,9	1239	21,1		
höherer Angestellter / Beamter	265	6,9	379	6,4		
selbständig	145	3,8	257	4,4		
Schüler / Student / Auszubildender / Umschüler	804	20,9	1231	20,9		
Rentner / Pensionär	401	10,4	621	10,6		
ohne Beruf / Hausfrau / -mann	782	20,3	1301	22,1		
Beschäftigung vor der Aufnahme					3,389	0,335
voll erwerbstätig	1248	32,5	1828	31,1		
teilzeitbeschäftigt	600	15,6	896	15,3		
nicht erwerbstätig	1100	28,7	1716	29,2		
arbeitslos	888	23,1	1432	24,4		
Berentung					0,434	0,510
ja	498	12,9	789	13,4		
nein	3348	87,1	5094	86,6		
Arbeitsfähigkeit					1,152	0,283
arbeits- / schulfähig teilweise arbeits- / schulfähig	668	27,0	975	25,8		
arbeits- / schul- / erwerbs- / berufsunfähig	1804	73,0	2804	74,2		

Die Teil- und die Gesamtstichprobe unterschieden sich hinsichtlich des Anteils der zehn am häufigsten dokumentierten Diagnosegruppen um maximal 4 % voneinander. Dies ergab ein Vergleich der einzelnen Diagnosegruppen der Teil- und der Gesamtstichprobe (s. Tabelle 13).

Tabelle 13: Vergleich der Teil- und der Gesamtstichprobe anhand der zehn häufigsten Diagnosen der Teilstichprobe

Diagnose	Teilstichprobe (n = 3862)		Gesamtstichprobe (n = 6481)		Differenz des Anteils (%)
	n	%	n	%	
F33	2717	70,4	4418	68,2	2,2
F60	1873	48,5	3045	47,0	1,5
F40	1786	46,2	2726	42,1	4,2
F43	1573	40,7	2485	38,3	2,4
F45	503	13,0	805	12,4	0,6
F50	489	12,7	801	12,4	0,3
F32	445	11,5	747	11,5	0,0
F44	376	9,7	585	9,0	0,7
F42	338	8,8	571	8,8	-0,1
F41	329	8,5	585	9,0	-0,5

Die Betrachtung des Anteils an Patienten, die die Therapie gegen ärztlichen Rat beendeten, ergab folgende Ergebnisse: Unter allen 6481 Patienten beendeten insgesamt 509 Patienten, entsprechend 7,9 % der Patienten, die Therapie gegen ärztlichen Rat. In der Teilstichprobe (n = 3862) brachen 26 Patienten (0,7 %) die Behandlung gegen ärztlichen Rat ab. Von den 1361 Patienten der Gesamtstichprobe mit einer Liegedauer von unter 42 Tagen beendeten 358 (26,3 %) die Behandlung gegen ärztlichen Rat.

3.2 Häufigkeit von Nonresponse und Verschlechterung

Der Anteil an Nonresponse und Verschlechterung variierte zwischen den Erhebungsarten. Die Ergebnisgruppen der indirekten Veränderungsmessung wurden, wie in Kapitel 1.3.2 dargestellt, gebildet. Der niedrigste gemessene Anteil an verschlechterten Patienten lag bei 1,1 %, gemessen mithilfe des BSS. Der höchste Anteil ergab sich bei der Messung mithilfe der SCL-90-R mit 5,9 % der Patienten. Die Anteile an Nonresponse waren in der direkten Messung aus Therapeutensicht mit 10,1 % am niedrigsten und mit 50,0 % bei der indirekten Messung mit dem IIP am höchsten. Im Folgenden sind die einzelnen Ergebnisse der direkten und indirekten Veränderungsmessung dargestellt.

3.2.1 Direkte Veränderungsmessung

Im Rahmen der direkten Veränderungsmessung gaben bei der Entlassung 286 von 3804 Patienten (7,5 %) an, dass sich ihre Symptomatik verschlechtert habe. Keine Veränderung ihrer Symptomatik bemerkten 481 (12,6 %) von ihnen. Aus Sicht der Therapeuten verschlechterte sich die Symptomatik bei 45 von 2204 Patienten (2,0 %) und blieb unverändert bei 222 (10,1 %) (s. Abbildung 4). Die Therapeuten schätzten die Veränderung der Symptomatik im Mittel signifikant besser ein als die Patienten selbst ($M_{Therapeut} = 4{,}14$, $SD = 0{,}69$, $M_{Patient} = 4{,}04$, $SD = 0{,}95$, $t = 4{,}740$, $p < 0{,}001$***).

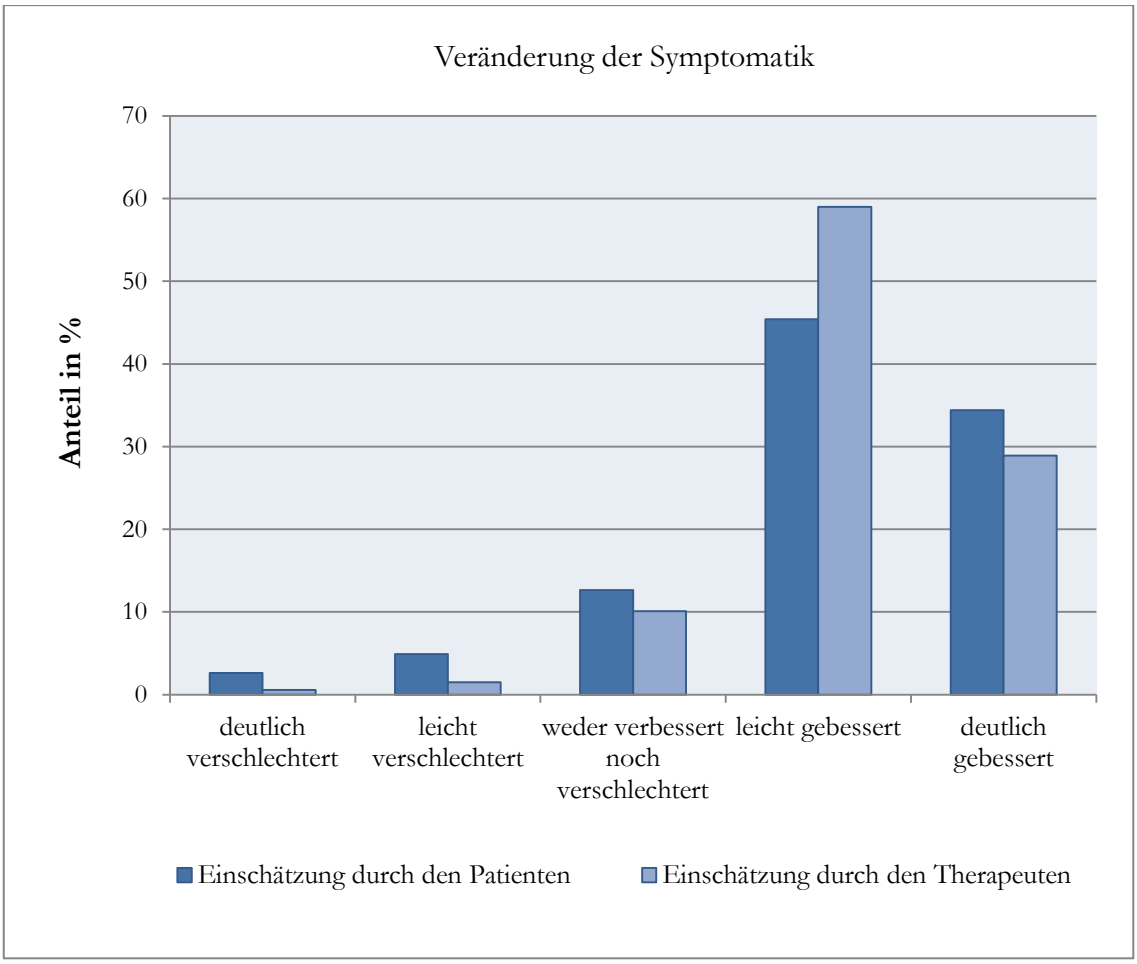

Abbildung 4: Beurteilung des Behandlungserfolgs zum Entlassungszeitpunkt anhand der Veränderung der Symptomatik aus Patienten- und Therapeutenperspektive

3.2.2 Indirekte Veränderungsmessung

Im Rahmen der indirekten Veränderungsmessung verschlechterten sich, gemessen mit dem GSI der SCL-90-R, 5,9 % der Patienten (229 von 3862 Patienten), gemessen mit dem IIP 5,0 % der Patienten (189 von 3814 Patienten). Aus Sicht der Therapeuten verschlechterten sich 23 von 2148 Patienten (1,1 %) gemessen mithilfe des BSS. Während sich die Anteile an Nonresponse in den Messungen mithilfe der SCL-90-R und des BSS ähnelten (23,4 % (904 von 3862) vs. 22,9 % (491 von 2148)), waren diese mit einem Anteil von 50,0 % (1906 von 3814 Patienten) im IIP deutlich höher (s. Abbildung 5).

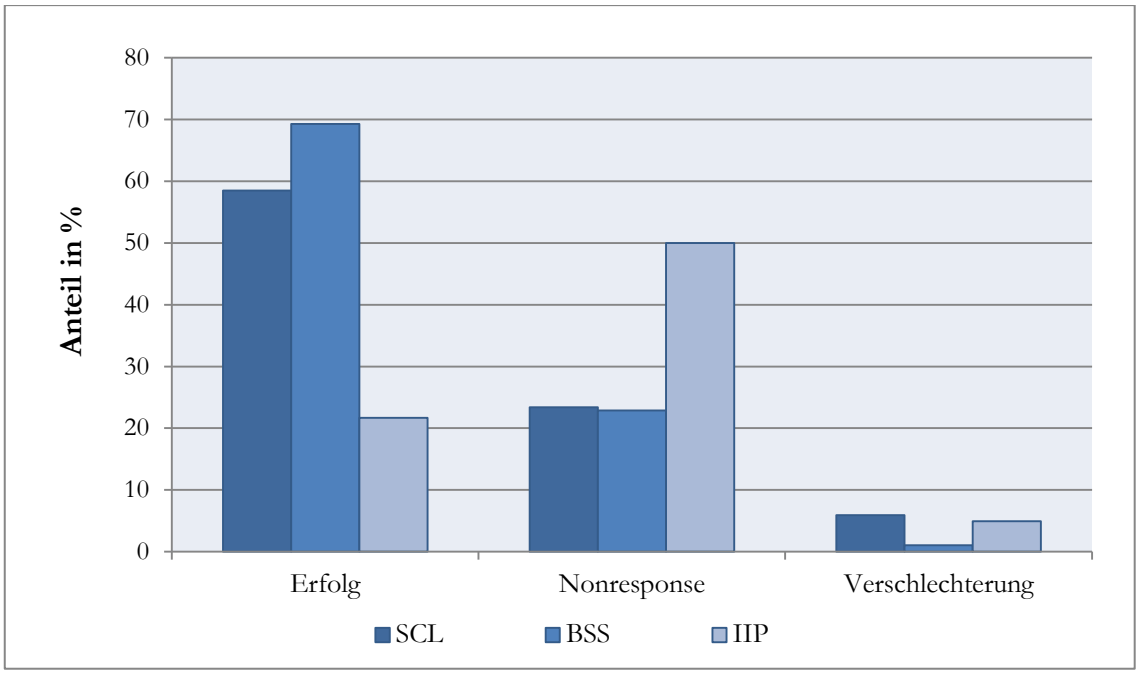

Abbildung 5: Behandlungsergebnisse (Erfolg, Nonresponse, Verschlechterung) im Rahmen der indirekten Veränderungsmessung, ohne Darstellung der Pseudogesunden

Die genauer differenzierten Behandlungsergebnisse sind in Abbildung 6 und Tabelle 14 dargestellt.

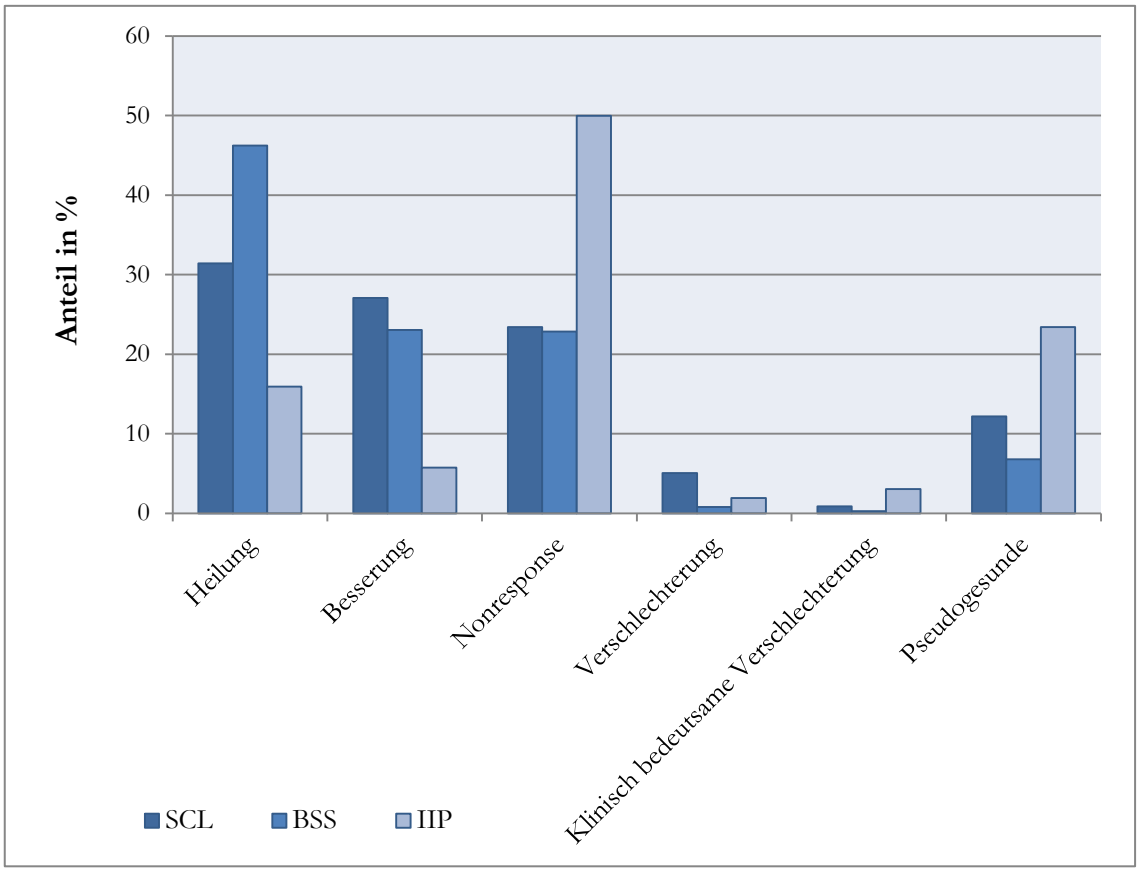

Abbildung 6: Behandlungsergebnisse im Rahmen der indirekten Veränderungsmessung

Tabelle 14: Behandlungsergebnisse im Rahmen der indirekten Veränderungsmessung

	SCL-90-R		BSS		IIP	
	n	%	n	%	n	%
Heilung	1213	31,4	993	46,2	607	15,9
Besserung	1046	27,1	495	23,0	219	5,7
Nonresponse	904	23,4	491	22,9	1906	50,0
Statistisch signifikante Verschlechterung	195	5,0	17	0,8	73	1,9
Klinisch bedeutsame Verschlechterung	34	0,9	6	0,3	116	3,0
Pseudogesunde	470	12,2	146	6,8	893	23,4

Die durchschnittliche Symptombelastung, erfasst mithilfe des GSI der SCL-90-R, nahm über die gesamte Stichprobe hinweg signifikant von 1,43 (SD = 0,65) bei Aufnahme auf 0,94 (SD = 0,67) bei Entlassung ab ($t = 53{,}47$, $p < 0{,}001$***). Die Effektstärke betrug für den GSI in diesem Fall $d = 0{,}75$, was nach Cohen (1988) einen mittleren Effekt darstellt. Die durchschnittliche Beeinträchtigungsschwere, gemessen mithilfe des BSS-Gesamtwerts, sank signifikant von 6,91 (SD = 1,76) bei Aufnahme auf 4,32 (SD = 1,75) bei Entlassung ($t = 63{,}19$, $p < 0{,}001$***), was mit einer Effektstärke von $d = 1{,}47$ einem hohen Effekt (Cohen 1988) entspricht. Die durchschnittliche Stärke interpersonaler Probleme, die sich im 64-Item-Gesamtwert des IIP widerspiegelte, reduzierte sich signifikant von 1,75 (SD = 0,52) bei Aufnahme auf 1,54 (SD = 0,59) zum Entlassungszeitpunkt ($t = 27{,}71$, $p < 0{,}001$***). Dieses kommt bei einer Effektstärke von $d = 0{,}41$ einem kleinen Effekt (Cohen 1988) gleich.

Zusammenfassend sind die Häufigkeiten von Nonresponse und Verschlechterung für die direkte und indirekte Veränderungsmessung in Tabelle 15 dargestellt.

Tabelle 15: Häufigkeit von Nonresponse und Verschlechterung in Abhängigkeit von der Messmethode

Veränderungs-messung	Fragebogen und Perspektive	Nonresponse (%)	Verschlechterung (%)
direkt	Patient	12,6	7,5
	Therapeut	10,1	2,0
indirekt	SCL, Patient	23,4	5,9
	BSS, Therapeut	22,9	1,1
	IIP, Patient	50,0	5,0

3.3 Nonresponse und Verschlechterung im Geschlechtervergleich

Im Folgenden werden die Ergebnisse der zweiten Forschungsfrage dargestellt. Zuerst wird auf die Ergebnisse der direkten und im Anschluss auf jene der indirekten Veränderungsmessung eingegangen.

3.3.1 Direkte Veränderungsmessung

Im Rahmen der direkten Veränderungsmessung ließen sich aus Sicht der Patienten und der Therapeuten bzgl. der Veränderung der Symptomatik keine signifikanten Geschlechtsunterschiede feststellen (s. Tabelle 16). Die Mittelwerte der Männer und der Frauen waren bei der Selbsteinschätzung der Veränderung der Symptomatik identisch (s. Tabelle 16). Ein Wert von 4 entsprach einer leichten Verbesserung der Symptomatik (s. 2.3.4). Die Therapeuten schätzten die Veränderung der Symptomatik sowohl ihrer männlichen als auch weiblichen Patienten im Mittel signifikant besser ein als diese selbst (für Frauen: $t = 3{,}76$, $p < 0{,}001$***; für Männer: $t = 2{,}89$, $p = 0{,}004$**).

Tabelle 16: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach dem Geschlecht

Perspektive	Geschlecht	n	M	SD	t	df [1]	p
Therapeuten-perspektive	männlich	817	4,14	0,66	-0,320	2202	0,749
	weiblich	1387	4,15	0,71			
Patienten-perspektive	männlich	1427	4,04	0,93	-0,124	3802	0,901
	weiblich	2377	4,04	0,96			

[1] df: Freiheitsgrade

Abbildung 7 und Abbildung 8 zeigen die Ergebnisse der direkten Veränderungsmessung der Symptomatik.

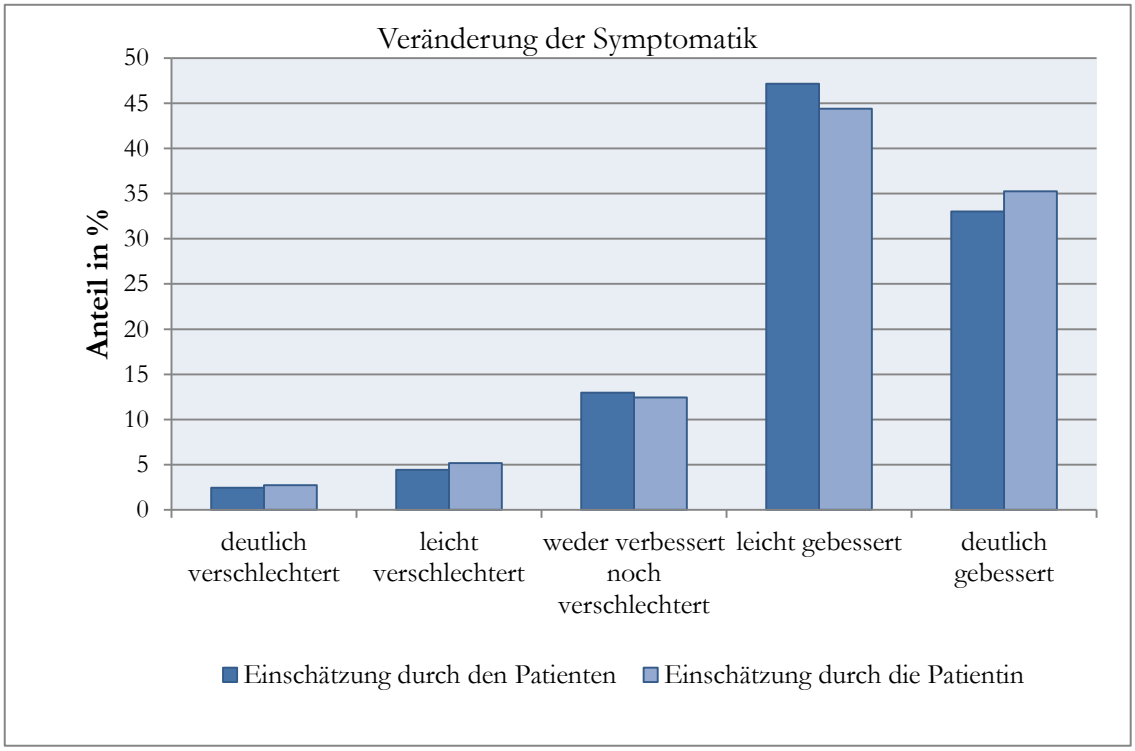

Abbildung 7: Selbsteinschätzung der Veränderung der Symptomatik differenziert nach dem Geschlecht

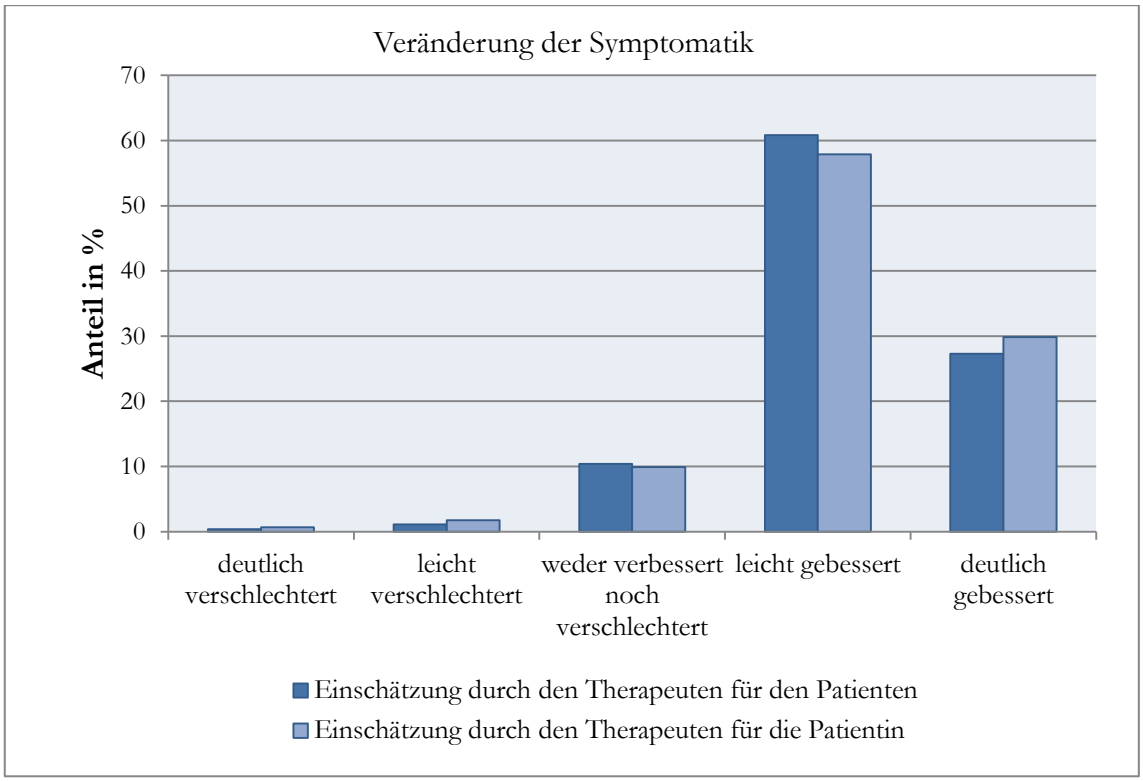

Abbildung 8: Therapeuteneinschätzung der Veränderung der Symptomatik differenziert nach dem Geschlecht

3.3.2 Indirekte Veränderungsmessung

Sowohl bei der zweistufigen Differenzierung Misserfolg versus Erfolg als auch bei der dreistufigen Unterscheidung Verschlechterung versus Nonresponse versus Erfolg ließen sich signifikante Geschlechtsunterschiede nur bei der Messung anhand des GSI der SCL-90-R feststellen (für Misserfolg vs. Erfolg: χ^2= 6,762, p = 0,010*, für Verschlechterung vs. Nonresponse vs. Erfolg χ^2 = 6,846, p = 0,033*), nicht jedoch bei Messungen, die auf dem BSS oder dem IIP basierten (s. Tabelle 17 und Tabelle 18).

Tabelle 17: Anzahl und Anteil der Patienten mit Erfolg und Misserfolg (inkl. Nonresponse) differenziert nach dem Geschlecht sowie Ergebnisse des χ^2-Tests

		Männer		Frauen		Statistik	
		n	%	n	%	χ^2	p
SCL	Erfolg	767	63,8	1492	68,2	6,762	0,010
	Misserfolg	436	36,2	697	31,8		
BSS	Erfolg	556	74,8	932	74,0	0,159	0,711
	Misserfolg	187	25,2	327	26,0		
IIP	Erfolg	282	26,9	544	29,1	1,571	0,215
	Misserfolg	767	73,1	1328	70,9		

Tabelle 18: Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Geschlecht sowie Ergebnisse des χ^2-Tests

		Männer		Frauen		Statistik	
		n	%	n	%	χ^2	p
SCL	Erfolg	767	63,8	1492	68,2	6,846	0,033
	Nonresponse	346	28,8	558	25,5		
	Verschlechterung	90	7,5	139	6,3		
BSS	Erfolg	556	74,8	932	74,0	0,237	0,888
	Nonresponse	178	24,0	313	24,9		
	Verschlechterung	9	1,2	14	1,1		
IIP	Erfolg	282	26,9	544	29,1	2,741	0,256
	Nonresponse	691	65,9	1215	64,9		
	Verschlechterung	76	7,2	113	6,0		

Deskriptiv lässt sich ableiten, dass Männer häufiger Misserfolge im Sinne von Verschlechterungen oder Nonresponse gemessen mit der SCL-90-R oder dem IIP aufwiesen als Frauen (s. Tabelle 17 und Tabelle 18). Bei der Differenzierung zwischen Erfolg und Misserfolg

mittels der SCL-90-R lag das standardisierte Residuum für Männer und Misserfolg bei 1,7 und für Frauen und Misserfolg bei -1,3. Wenn zusätzlich die Gruppe Nonresponse unterschieden wurde, betrug das standardisierte Residuum für Nonresponse und Männer 1,4 und für Nonresponse und Frauen -1,1. Männer wiesen also häufiger und Frauen seltener Misserfolge oder Nonresponse auf, als bei statistischer Unabhängigkeit zu erwarten wäre. Die Beträge der standardisierten Residuen waren jedoch < 2, sodass die Unterschiede nicht als signifikant bewertet werden können. Cramers V betrug für die SCL-90-R V = 0,045. Wenn der Betrag eines Zusammenhangsmaßes, in diesem Fall Cramers V, kleiner als 0,1 ist, besteht kein Zusammenhang zwischen den beiden untersuchten Variablen (Kuckartz et al. 2013). Cramers V bezieht sich dabei auf die Stärke des Zusammenhangs in der Stichprobe, nicht auf den Zusammenhang in der Grundgesamtheit (Duller 2013).

In Ergänzung zur Untersuchung dieser Ergebnisgruppen wurden die Summenscores der einzelnen Testverfahren im Rahmen einer Varianzanalyse auf Unterschiede zwischen Männern und Frauen untersucht. Die Analyse der Summenscores der Tests ergab für alle drei Testverfahren der indirekten Veränderungsmessung einen signifikanten Haupteffekt für die Gruppenzugehörigkeit (Männer vs. Frauen) (SCL-90-R: $F(1,3860) = 80,470$, $p < 0,001$***, BSS: $F(1, 2146) = 4,884$, $p < 0,027$*, IIP: $F(1, 3812) = 6,661$, $p = 0,010$*) (s.Tabelle 19).

Tabelle 19: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Geschlechts

		Männer			Frauen			Statistik	
		M	SD	d	M	SD	d	Haupteffekt Gruppen-zugehörigkeit	
								F	p
GSI	Aufnahme	1,30	0,63	0,71	1,50	0,65	0,79	80,47	< 0,001
	Entlassung	0,85	0,64		0,99	0,68			
BSS-Gesamt-wert	Aufnahme	6,90	1,72	1,57	6,99	1,81	1,43	4,88	0,027
	Entlassung	4,20	1,65		4,40	1,81			
IIP $_{ges}$	Aufnahme	1,73	0,54	0,39	1,77	0,51	0,42	6,66	0,010
	Entlassung	1,52	0,62		1,56	0,58			

Deskriptiv zeigte sich, dass die Frauen unabhängig vom Erhebungszeitpunkt höhere Werte erreichten. Dies bedeutet die Angabe einer höheren Symptombelastung und stärkerer interpersonaler Probleme aus Patientenperspektive sowie einer stärkeren Belastung aus Therapeutenperspektive (s. Tabelle 19 und Abbildung 9). Die Effektstärke der Gruppenzugehörigkeit zu den Männern oder Frauen lag für den BSS und das IIP bei $\eta_{partial}^2 = 0,002$ und für die SCL-90-R bei $\eta_{partial}^2 = 0,020$ bei sehr hoher Teststärke (beobachtete Schärfe =

1,00). Ein kleiner Effekt nach der Definition von Cohen (1988) lag somit nur für die Messung mit der SCL-90-R vor. Die Unterscheidung zwischen Männern und Frauen erklärte also 2 % der Variation des GSI. Die Wechselwirkungsdiagramme sind in Abbildung 9 dargestellt.

Der Haupteffekt des Messzeitpunktes war für alle drei Tests signifikant (SCL-90-R: $F(1, 3860) = 2595{,}048$, $p < 0{,}001$***, BSS: $F(1, 2146) = 3770{,}112$, $p < 0{,}001$***, IIP: $F(1, 3812) = 711{,}791$, $p < 0{,}001$***) und wies jeweils eine gute Teststärke auf (beobachtete Schärfe = 1,00). Deskriptiv zeigten sich nach der Therapie unabhängig vom Geschlecht niedrigere Summenscores (s. Tabelle 19 und Abbildung 9). Bei der Messung mit der SCL-90-R war die Effektstärke für den Zeitfaktor mit $\eta^2 = 0{,}40$ groß. Somit konnten 40 % der Variation des GSI einer Person, also der generellen psychischen Belastung, durch den Messzeitpunkt erklärt werden. Beim IIP ließen sich 15,7 % ($\eta^2 = 0{,}157$) und beim BSS 63,7 % ($\eta^2 = 0{,}637$) der Variation der Gesamtwerte durch den Messzeitpunkt erklären. Auch diese beiden Effektstärken entsprechen einem großen Effekt nach Cohen (1988).

Die Wechselwirkung zwischen dem Geschlecht und dem Messzeitpunkt stellte sich nur bei der Messung mithilfe der SCL-90-R signifikant dar (SCL-90-R: $F(1,3860) = 10{,}924$, $p = 0{,}001$**, BSS: $F(1, 2146) = 1{,}508$, $p = 0{,}220$, n. s., IIP: $F(1, 3812) = 0{,}150$, $p = 0{,}699$, n. s.). Diese Wechselwirkung bei der Messung mit der SCL-90-R erklärt jedoch bei guter Teststärke (beobachtete Schärfe = 0,91) nur 0,2 % der Variation des GSI einer Person ($\eta^2 = 0{,}002$).

Für alle drei Tests (SCL-90-R, BSS und IIP) ergab der Mauchly-Test auf Sphärizität, dass die Voraussetzungen für eine zweifaktorielle Varianzanalyse mit Messwiederholung – homogene Messzeitpunktvarianzen und Kovarianzen (Bühner und Ziegler 2009) – erfüllt waren.

Der Levene-Test auf Gleichheit der Fehlervarianzen zeigte, dass die Gruppenvarianzen bei Männern und Frauen bei der Messung mit der SCL-90-R ($F(1,3860) = 10{,}577$, $p = 0{,}001$**) und dem BSS ($F(1, 2146) = 14{,}573$, $p < 0{,}001$***) zum Entlassungszeitpunkt sowie zu beiden Zeitpunkten bei der Messung mit dem IIP (bei Aufnahme $F(1, 3812) = 8{,}066$, $p = 0{,}005$**; bei Entlassung $F(1, 3812) = 8{,}493$, $p = 0{,}004$**) nicht homogen waren. Da das Verhältnis der Anzahl von Frauen zu Männern mit 1,7 zu 1 kleiner als 4:1 war, durfte der Wert des F_{max}-Tests maximal 10 betragen (Bühner und Ziegler 2009). Das Verhältnis der Varianzen der beiden Gruppen war mit 1,14 für die SCL-90-R-Messung, 1,21 für die BSS-Messung, 1,13 für die IIP-Messung bei der Aufnahme und 1,16 für die IIP-Messung bei der Entlassung jeweils kleiner als 10, sodass das α-Niveau für keinen der Tests nach unten korrigiert werden musste.

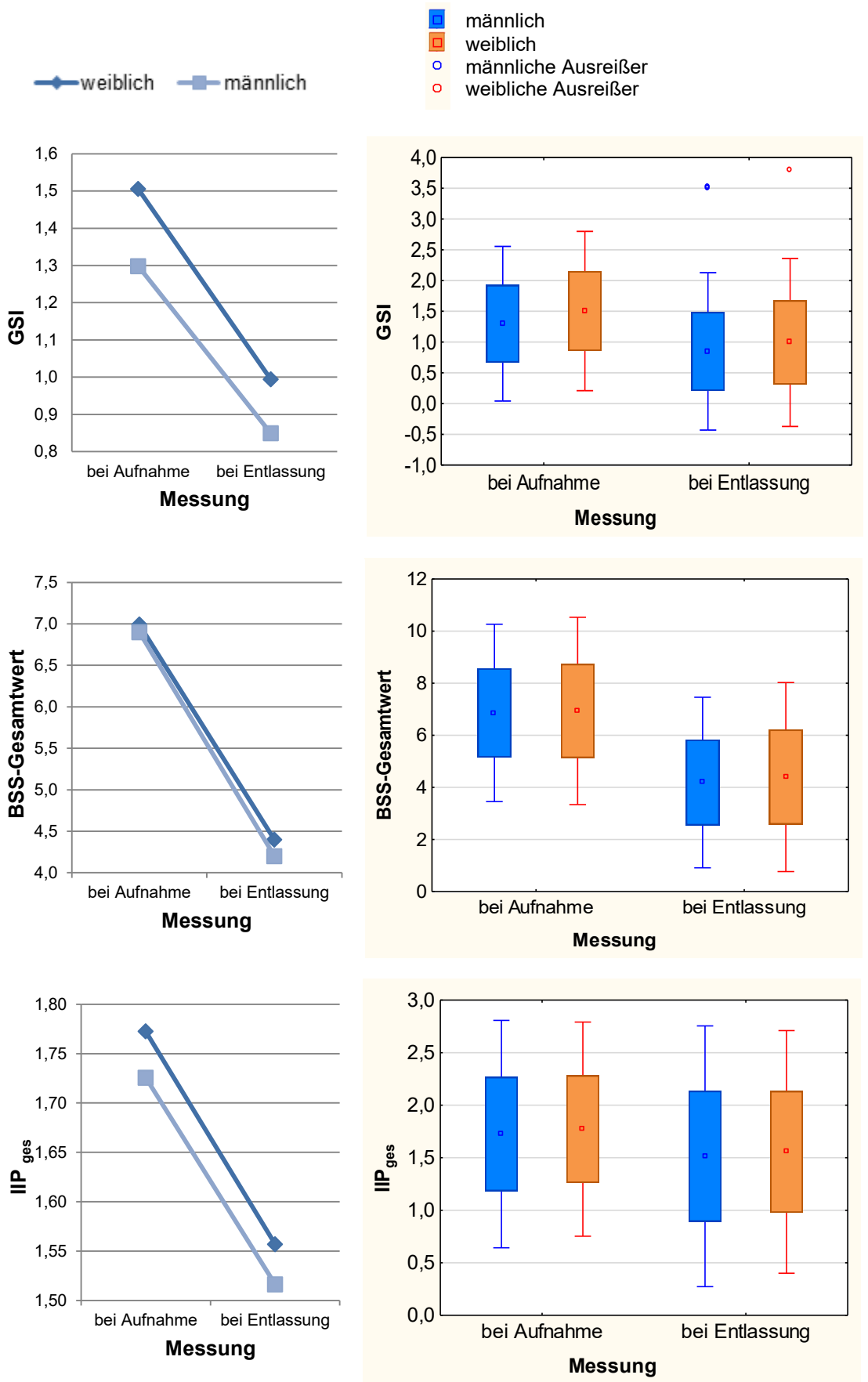

Abbildung 9: Summenscores differenziert nach dem Geschlecht und Messzeitpunkt (links: Mittelwerte; rechts: Mittelwerte, 1 SD, 2 SD; oben: SCL-90-R; mittig: BSS; unten: IIP)

3.4 Nonresponse und Verschlechterungen bei Persönlichkeitsstörungen

Von den 3862 untersuchten Patienten wiesen 66 (1,7 %) eine Cluster-A-Persönlichkeitsstörung, 896 (23,2 %) eine Cluster-B- und 884 (22,9 %) eine Cluster-C-Persönlichkeitsstörung auf. Sieben Patienten mit Persönlichkeitsstörungen konnten keinem der Cluster zugeordnet werden. Insgesamt wurde bei 1853 Patienten eine Persönlichkeitsstörung diagnostiziert (48,0 % der Patienten).

3.4.1 Ergebnisse für Personen mit und ohne Persönlichkeitsstörungen

In diesem Kapitel wird beschrieben, wie sich Personen mit und ohne Persönlichkeitsstörungen in der direkten und indirekten Veränderungsmessung unterschieden.

3.4.1.1 Direkte Veränderungsmessung

Die Veränderung der Symptomatik von Patienten ohne Persönlichkeitsstörungen wurde von Patienten und Therapeuten im Mittel signifikant positiver (höhere Skalenwerte) als die Veränderung der Symptomatik von Patienten mit Persönlichkeitsstörungen eingeschätzt (s. Tabelle 20). Deskriptiv zeigte sich, dass die Therapeuten dabei die Veränderung der Symptomatik der Patienten im Mittel positiver einschätzten als die Patienten selbst (s. Tabelle 20).

Tabelle 20: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach dem Vorliegen von Persönlichkeitsstörungen

Perspektive	Persönlichkeitsstörung	n	M	SD	t	df	p
Therapeutenperspektive	nein	1214	4,17	0,688	2,451	2107	0,014
	ja	990	4,10	0,694			
Patientenperspektive	nein	1987	4,10	0,933	3,716	3802	< 0,001
	ja	1817	3,98	0,963			

Es schätzten 8,2 % der Patienten mit Persönlichkeitsstörungen und 6,9 % der Patienten ohne Persönlichkeitsstörung ihre Symptomatik als verschlechtert ein. Die Therapeuten beurteilten die Symptomatik von 2,4 % der Patienten mit Persönlichkeitsstörungen und von 1,7 % der Patienten ohne Persönlichkeitsstörungen als verschlechtert. Eine unveränderte Symptomatik beschrieben 13,3 % der Patienten mit Persönlichkeitsstörungen und 12,0 % der Patienten ohne Persönlichkeitsstörungen. Die Therapeuten stuften die Symptomatik von 10,4 % der Patienten mit und 9,8 % der Patienten ohne Persönlichkeitsstörungen als unverändert ein (s. Abbildung 10 und Abbildung 11).

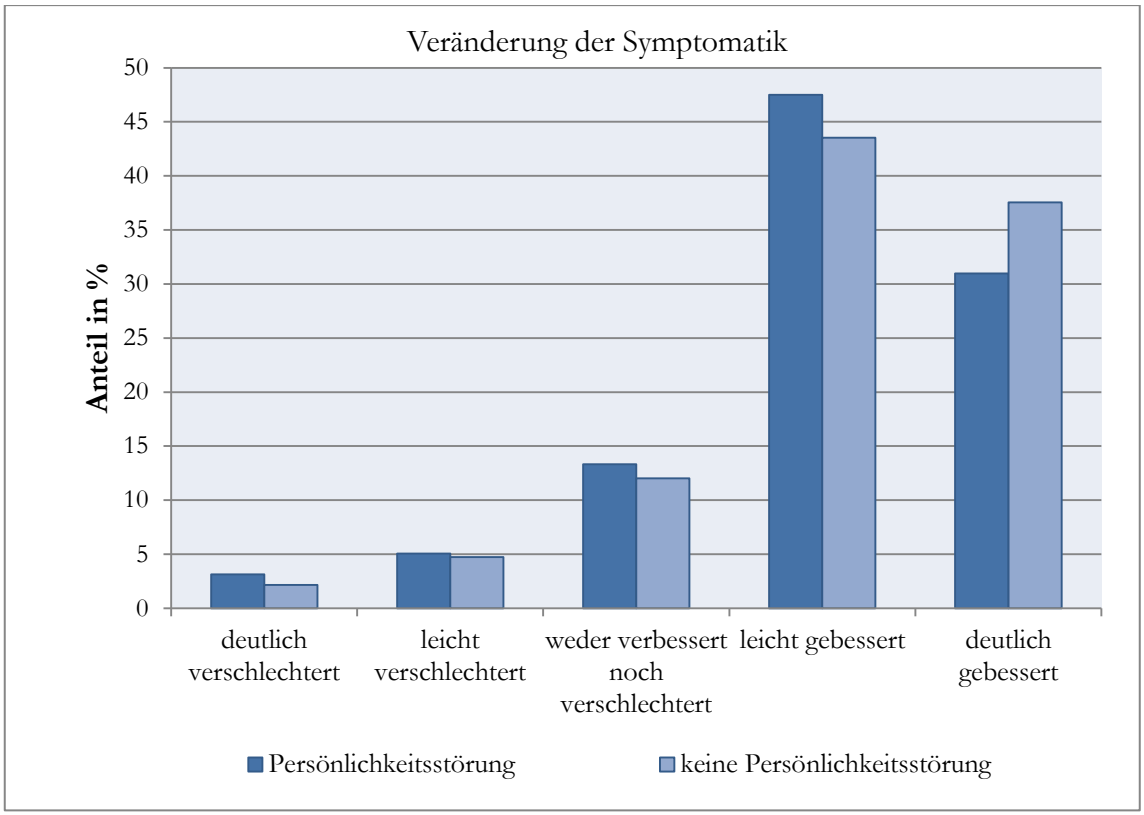

Abbildung 10: Veränderung der Symptomatik aus Patientenperspektive differenziert nach dem Vorliegen einer Persönlichkeitsstörung

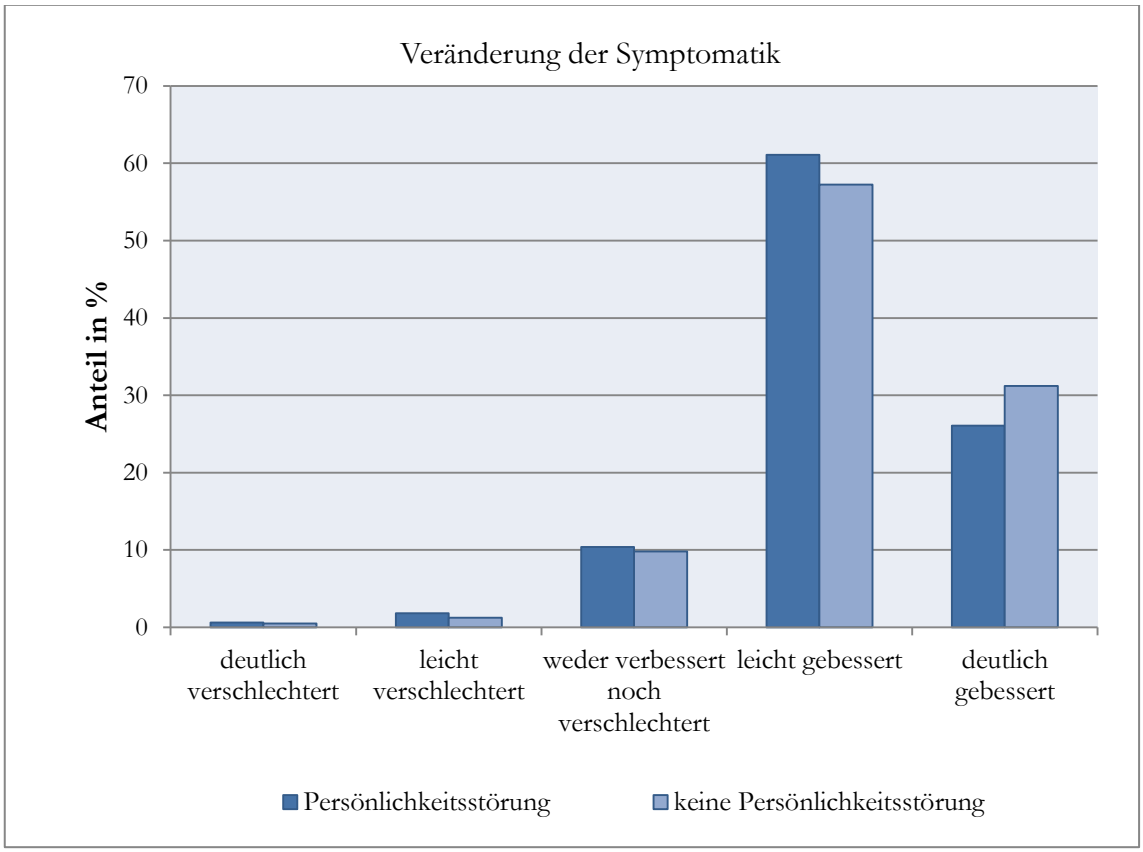

Abbildung 11: Veränderung der Symptomatik aus Therapeutenperspektive differenziert nach dem Vorliegen einer Persönlichkeitsstörung

3.4.1.2 Indirekte Veränderungsmessung

Patienten mit und ohne Persönlichkeitsstörung unterschieden sich in der indirekten Veränderungsmessung nicht signifikant im Hinblick auf die Zuordnung zu den Gruppen der Verschlechterten, der Nonresponder und der erfolgreich Therapierten bzw. zu den Gruppen Erfolg vs. Misserfolg (s. Tabelle 21 und Tabelle 22). Dies ergab die Analyse der Ergebnisse aller drei Testverfahren (SCL-90-R, BSS und IIP).

Tabelle 21: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Erfolge und Misserfolge (inkl. Nonresponse) differenziert nach dem Vorliegen einer Persönlichkeitsstörung

		Patienten ohne Persönlichkeitsstörungen		Patienten mit Persönlichkeitsstörungen		Statistik	
		n	%	n	%	χ^2	P
SCL	Erfolg	1172	67,5	1087	65,7	1,225	0,275
	Misserfolg	565	32,5	568	34,3		
BSS	Erfolg	825	75,0	663	73,5	0,582	0,472
	Misserfolg	275	25,0	239	26,5		
IIP	Erfolg	415	28,7	411	27,9	0,251	0,622
	Misserfolg	1031	71,3	1064	72,1		

Tabelle 22: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Vorliegen einer Persönlichkeitsstörung

		Patienten ohne Persönlichkeitsstörungen		Patienten mit Persönlichkeitsstörungen		Statistik	
		n	%	n	%	χ^2	p
SCL	Erfolg	1172	67,5	1087	65,7	1,226	0,541
	Nonresponse	451	26,0	453	27,4		
	Verschlechterung	114	6,6	115	6,9		
BSS	Erfolg	825	75,0	663	73,5	1,110	0,585
	Nonresponse	261	23,7	230	25,5		
	Verschlechterung	14	1,3	9	1,0		
IIP	Erfolg	415	28,7	411	27,9	2,572	0,274
	Nonresponse	928	64,2	978	66,3		
	Verschlechterung	103	7,1	86	5,8		

Die Varianzanalysen, die den Messzeitpunkt und die Gruppenzugehörigkeit als Haupteffekte sowie deren Wechselwirkung untersuchten, zeigten für alle drei Testverfahren einen signifikanten Haupteffekt für die Gruppenzugehörigkeit (Patienten mit Persönlichkeitsstörung vs. Patienten ohne Persönlichkeitsstörung) (SCL-90-R: $F(1, 3860) = 33{,}401$, $p < 0{,}001$***, $\eta_{partial}^2 = 0{,}009$; BSS: $F(1, 2146) = 11{,}922$, $p = 0{,}001$**, $\eta_{partial}^2 = 0{,}006$; IIP: $F(1,3812) = 62{,}119$, $p < 0{,}001$***, $\eta_{partial}^2 = 0{,}016$). Personen mit Persönlichkeitsstörungen bzw. deren Therapeuten gaben im Vergleich zu Personen ohne Persönlichkeitsstörungen und deren Therapeuten unabhängig vom Messzeitpunkt eine höhere Symptombelastung, eine schwerere Beeinträchtigung durch die Erkrankung und stärkere interpersonale Probleme an (s. Tabelle 23 und Abbildung 12). 1,6 % der Variation im IIP-Gesamtwert zwischen den Personen ließ sich durch das Vorliegen einer Persönlichkeitsstörung erklären, was einem kleinen Effekt nach Cohen (1988) entspricht. Bei der Erhebung mit der SCL-90-R ließen sich 0,9 % und bei der Messung mit dem BSS 0,6 % der Variation der Skalensummenwerte zwischen den Personen durch die Gruppenzugehörigkeit erklären. Dieses genügt Cohen (1988) zufolge noch nicht, um von einem kleinen Effekt sprechen zu können.

Tabelle 23: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Vorliegens einer Persönlichkeitsstörung

		Patienten ohne Persönlichkeitsstörungen			Patienten mit Persönlichkeitsstörungen			Statistik	
		M	SD	d	M	SD	d	Haupteffekt Gruppenzugehörigkeit	
								F	P
GSI	Aufnahme	1,38	0,63	0,78	1,48	0,66	0,74	33,401	< 0,001
	Entlassung	0,89	0,65		1,00	0,69			
BSS-Gesamtwert	Aufnahme	6,88	1,77	1,51	7,05	1,78	1,45	11,922	0,001
	Entlassung	4,21	1,75		4,47	1,75			
IIP $_{ges}$	Aufnahme	1,69	0,53	0,39	1,83	0,51	0,44	62,119	< 0,001
	Entlassung	1,49	0,58		1,60	0,60			

Es lag für alle drei Tests der indirekten Veränderungsmessung ein signifikanter Haupteffekt des Erhebungszeitpunkts vor (SCL-90-R: $F(1, 3860) = 2852{,}550$, $p < 0{,}001$, $\eta^2 = 0{,}425$; BSS: $F(1, 2146) = 3941{,}466$, $p < 0{,}001$***, $\eta^2 = 0{,}647$; IIP: $F(1,3812) = 769{,}712$, $p < 0{,}001$***, $\eta^2 = 0{,}168$). Die Skalensummenwerte bzw. globalen Kennwerte waren bei der Entlassung niedriger als bei der Aufnahme (s. Tabelle 23 und Abbildung 12).

Der Wechselwirkungseffekt zwischen dem Vorliegen einer Persönlichkeitsstörung und dem Messzeitpunkt war für keinen der Tests signifikant (SCL-90-R: $F(1, 3860) = 0{,}123$, $p = 0{,}726$, n. s.; BSS: $F(1, 2146) = 1{,}138$, $p = 0{,}286$, n. s.; IIP: $F(1,3812) = 1{,}820$, $p = 0{,}177$, n. s.).

Die Gruppenvarianzen bei Patienten mit und ohne Persönlichkeitsstörungen waren, wie die Levene-Tests auf Gleichheit der Fehlervarianzen ergaben, bei der SCL-90-R zu beiden Zeitpunkten ($F(1,3860) = 4{,}264$, $p = 0{,}039^*$ bei Aufnahme und $F(1,3860) = 6{,}474$, $p = 0{,}011^*$ bei Entlassung) und beim IIP bei Aufnahme ($F(1,3812) = 4{,}676$, $p = 0{,}031^*$) inhomogen. F_{max} war in allen drei Fällen mit 1,1 kleiner als 10 bei einem Verhältnis von Personen ohne Persönlichkeitsstörungen zu Personen mit Persönlichkeitsstörungen von 1,1:1, sodass die Varianzanalysen ohne Korrektur des α-Niveaus ausgewertet werden konnten.

Homogene Messzeitpunktvarianzen und Kovarianzen als Voraussetzung für eine zweifaktorielle Varianzanalyse mit Messwiederholung waren für alle drei Tests – dem Mauchly-Test auf Sphärizität nach – gegeben.

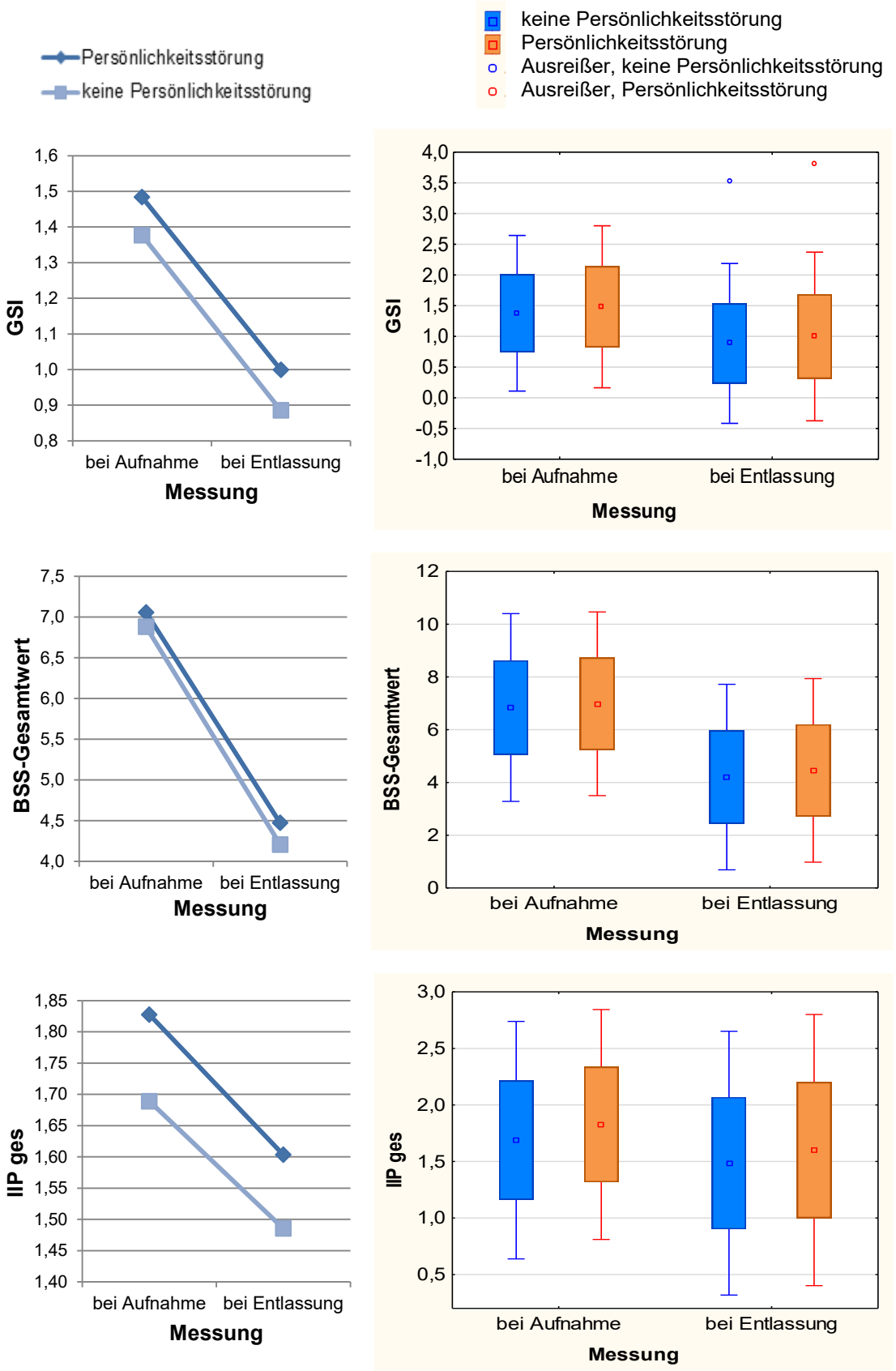

Abbildung 12: Mittelwerte der Summenscores differenziert nach dem Vorliegen von Persönlichkeitsstörungen und dem Messzeitpunkt (links: Mittelwerte; rechts: Mittelwerte, 1 SD, 2 SD; oben: SCL-90-R; mittig: BSS; unten: IIP)

3.4.2 Ergebnisse für Personen mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen

In diesem Kapitel werden die Therapieergebnisse der Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen verglichen. Unterschiede zu Personen mit Cluster-A-Persönlichkeitsstörungen wurden nicht untersucht, da ihr Anteil in der Stichprobe sehr gering war.

3.4.2.1 Direkte Veränderungsmessung

Die Gruppen der Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen unterschieden sich nicht signifikant in der Veränderung der Symptomatik voneinander (s. Tabelle 24).

Tabelle 24: Einschätzung der Veränderung der Symptomatik aus Therapeuten- und Patientenperspektive differenziert nach den Persönlichkeitsstörungsclustern B und C

Perspektive	Cluster	n	M	SD	t	df	p
Therapeuten-perspektive	B	465	4,09	0,73	-0,643	940	0,520
	C	477	4,12	0,67			
Patientenperspektive	B	871	3,95	1,03	-1,184	1719	0,237
	C	873	4,01	0,92			

Aus der Perspektive der Patienten verschlechterten sich 9,2 % (80 von 871) der Patienten mit Cluster-B- und 7,7 % (67 von 873) der Patienten mit Cluster-C-Persönlichkeitsstörung. 13,7 % der Patienten mit Cluster-B-Persönlichkeitsstörung (119 von 871) und 12,6 % der Patienten mit Cluster-C-Persönlichkeitsstörung (110 von 873) gaben eine unveränderte Symptomatik an, während 77,2 % (672 von 871) der Patienten mit Cluster-B- und 79,7 % (696 von 873) der Patienten mit Cluster-C-Persönlichkeitsstörung eine Verbesserung der Symptomatik nannten (s. Abbildung 13). Die Therapeuten schätzten die Symptomatik von 3,2 % (15 von 465) der Patienten mit Cluster-B-Persönlichkeitsstörung als verschlechtert, von 10,8 % (50 von 465) von ihnen als unverändert und von 86,0 % (400 von 465) von ihnen als verbessert ein. Die Symptomatik der Patienten mit Cluster-C-Persönlichkeitsstörung klassifizierten die Therapeuten zu 1,7 % (8 von 477) als verschlechtert, zu 10,1 % als unverändert (48 von 477) und zu 88,3 % (421 von 477) als verbessert (s. Abbildung 14)

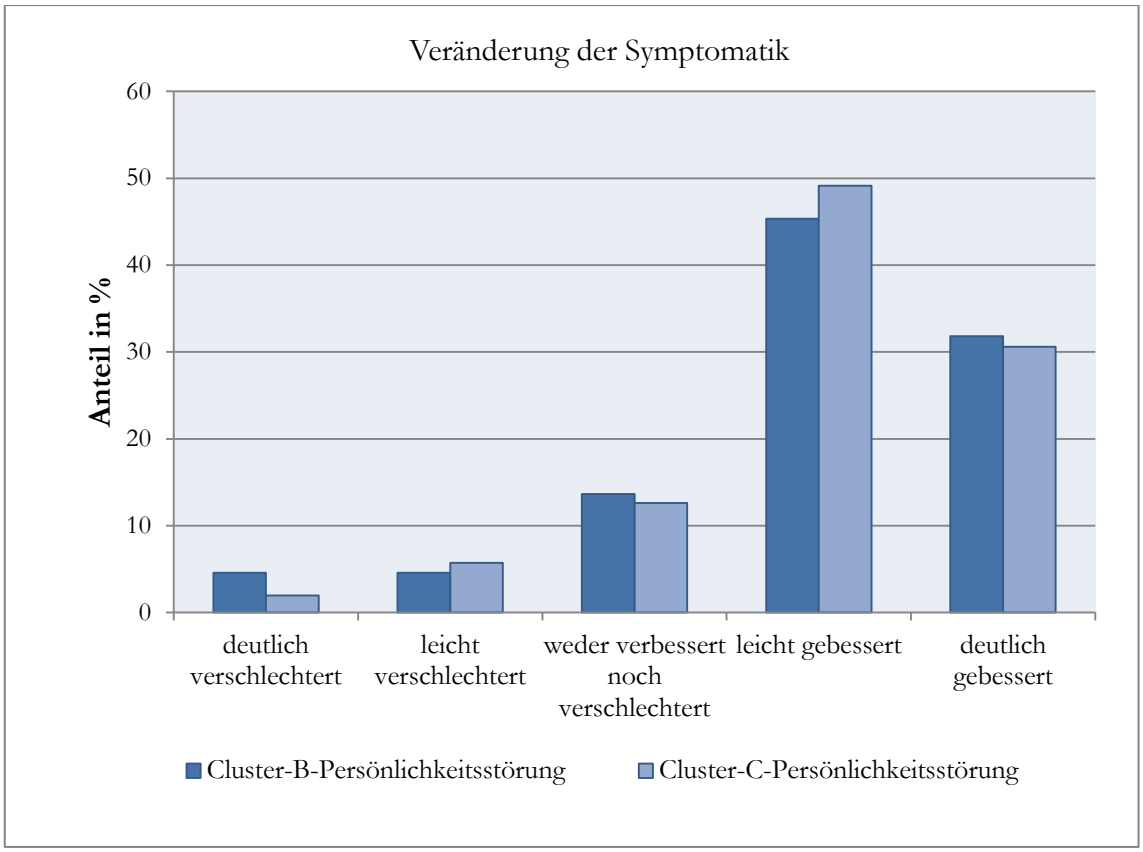

Abbildung 13: Veränderung der Symptomatik aus Patientenperspektive differenziert nach Persönlichkeitsstörungsclustern

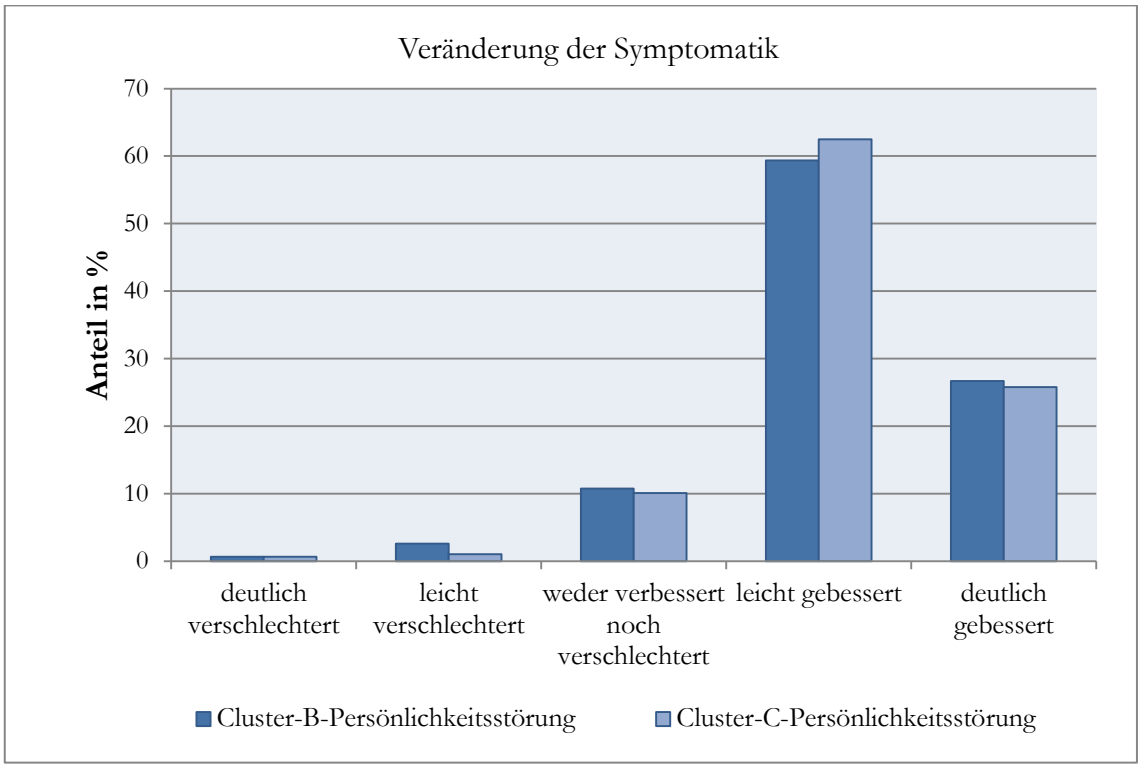

Abbildung 14: Veränderung der Symptomatik aus Therapeutenperspektive differenziert nach Persönlichkeitsstörungsclustern

3.4.2.2 Indirekte Veränderungsmessung

Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen wiesen keine signifikanten Unterschiede in der Häufigkeit von Erfolg vs. Misserfolg (inklusive Nonresponse) oder Erfolg vs. Nonresponse vs. Verschlechterung auf (s. Tabelle 25 und Tabelle 26).

Tabelle 25: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Erfolge und Misserfolge (inkl. Nonresponse) differenziert nach dem Vorliegen einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung

		Cluster B		Cluster C		Statistik	
		n	%	n	%	χ^2	p
SCL	Erfolg	529	65,1	523	67,0	0,586	0,459
	Misserfolg	283	34,9	258	33,0		
BSS	Erfolg	305	72,6	323	74,1	0,234	0,628
	Misserfolg	115	27,4	113	25,9		
IIP	Erfolg	210	29,7	189	26,8	1,499	0,237
	Misserfolg	497	70,3	517	73,2		

Tabelle 26: Ergebnisse des χ^2-Tests sowie Anzahl und Anteil der Patienten mit Erfolg, Nonresponse oder Verschlechterung differenziert nach dem Vorliegen einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung

		Cluster B		Cluster C		Statistik	
		n	%	n	%	χ^2	p
SCL	Erfolg	529	65,1	523	67,0	0,691	0,704
	Nonresponse	225	27,7	208	26,6		
	Verschlechterung	58	7,1	50	6,4		
BSS	Erfolg	305	72,6	323	74,1	1,331	0,512
	Nonresponse	112	26,7	107	24,5		
	Verschlechterung	3	0,7	6	1,4		
IIP	Erfolg	210	29,7	189	26,8	3,909	0,140
	Nonresponse	464	65,6	469	66,4		
	Verschlechterung	33	4,7	48	6,8		

Bei der Analyse der Summenscores der Personen mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen unterschieden sich diese hinsichtlich des Haupteffekts der Gruppenzugehörigkeit unabhängig vom Messzeitpunkt nur in der Messung mithilfe der SCL-90-R signifikant voneinander (SCL-90-R: $F(1,1778) = 24,956$, $p < 0,001$***, $\eta_{partiell}^2 = 0,014$).

Es ließen sich 1,4 % der Variation des GSI durch das vorliegende Persönlichkeitsstörungscluster erklären, was nach Cohen (1988) einem kleinen Effekt entspricht. Personen mit Cluster-B-Persönlichkeitsstörung wiesen bei Aufnahme und bei Entlassung eine höhere Symptombelastung auf als Patienten mit Cluster-C-Persönlichkeitsstörung (s. Tabelle 27 und Abbildung 15). Für den BSS und das IIP lag kein signifikanter Haupteffekt der Gruppenzugehörigkeit vor (BSS: $F(1,919) = 2,283$, $p = 0,131$, n. s.; IIP: $F(1,1745) = 0,268$, $p = 0,605$, n. s.). In der Beeinträchtigungsschwere aus Therapeutensicht und der Stärke interpersoneller Probleme unterschieden sich die Patienten der beiden Clustergruppen somit nicht signifikant (s. Tabelle 27 und Abbildung 15).

Tabelle 27: Effektstärken und zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor unter Berücksichtigung der Skalensummenwerte bei Aufnahme und Entlassung und des Vorliegens einer Cluster-B- oder Cluster-C-Persönlichkeitsstörung

		Cluster B			Cluster C			Statistik	
		M	SD	d	M	SD	d	Haupteffekt Gruppenzugehörigkeit	
								F	p
GSI	Aufnahme	1,56	0,68	0,72	1,42	0,63	0,77	24,956	< 0,001
	Entlassung	1,07	0,72		0,93	0,64			
BSS-Gesamt-Wert	Aufnahme	7,14	1,81	1,44	6,96	1,77	1,44	2,283	0,131
	Entlassung	4,52	1,79		4,40	1,75			
IIP_{ges}	Aufnahme	1,84	0,52	0,46	1,81	0,50	0,42	0,268	0,605
	Entlassung	1,60	0,62		1,60	0,59			

Die Skalensummenwerte waren bei der Entlassung niedriger als bei der Aufnahme (s. Tabelle 27 und Abbildung 15). Unabhängig vom Persönlichkeitsstörungscluster lag ein signifikanter, nach Cohen (1988) hoher Haupteffekt des Erhebungszeitpunktes vor (SCL-90-R: $F(1, 1778) = 1333,123$, $p < 0,001$***, $\eta^2 = 0,429$; BSS: $F(1, 919) = 1673,256$, $p < 0,001$***, $\eta^2 = 0,645$; IIP: $F(1, 1745) = 394,265$, $p < 0,001$***, $\eta^2 = 0,184$).

Die Wechselwirkung zwischen dem Persönlichkeitsstörungscluster und dem Messzeitpunkt war für keinen der Tests signifikant (SCL-90-R: $F(1, 1778) = 0,001$, $p = 0,978$ n. s.; BSS: $F(1, 919) = 0,201$, $p = 0,654$, n. s.; IIP: $F(1, 1745) = 2,288$, $p = 0,131$, n. s.).

Die Gruppenvarianzen bei Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen waren, wie die Levene-Tests auf Gleichheit der Fehlervarianzen ergaben, bei der SCL-90-R zu beiden Zeitpunkten ($F(1,1778) = 5,291$, $p = 0,022$* bei Aufnahme und $F(1,1778) = 19,893$, $p < 0,001$*** bei Entlassung) und beim IIP bei Entlassung ($F(1,1745) = 5,097$, $p = 0,024$*) inhomogen. F_{max} war in allen drei Fällen jedoch kleiner als

10 (SCL-90-R bei Aufnahme F_{max} = 1,15, bei Entlassung F_{max} = 1,26; IIP bei Entlassung F_{max} = 1,11) bei einem Verhältnis von Personen mit Cluster-B-Persönlichkeitsstörungen zu Personen mit Cluster-C-Persönlichkeitsstörungen von 1,1:1, sodass die Varianzanalysen ohne Korrektur des α-Niveaus ausgewertet werden konnten.

Dem Mauchly-Test auf Sphärizität nach waren homogene Messzeitpunktvarianzen und Kovarianzen als Voraussetzung für eine zweifaktorielle Varianzanalyse mit Messwiederholung für alle drei Tests gegeben.

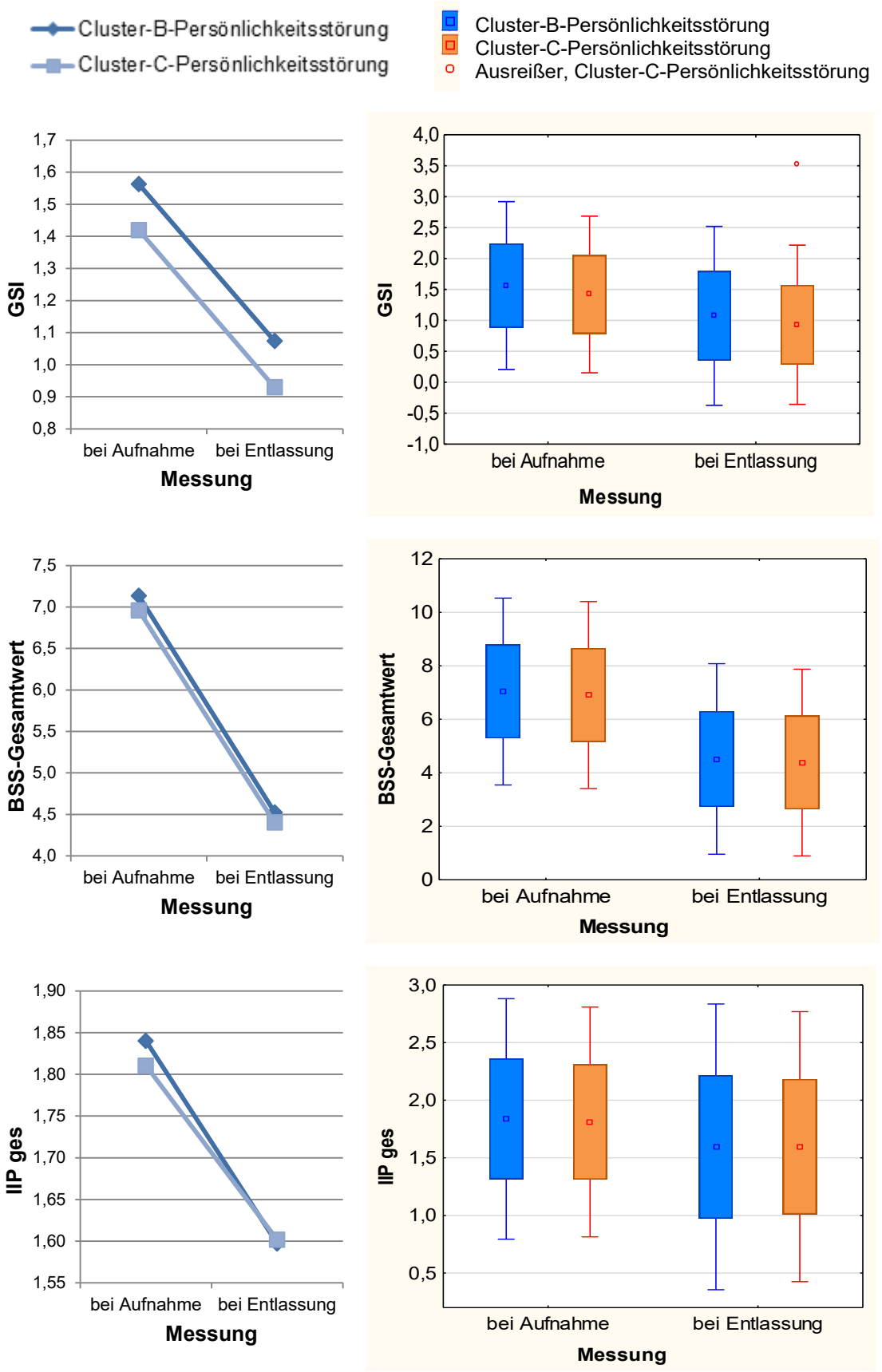

Abbildung 15: Summenscores differenziert nach dem Persönlichkeitsstörungscluster und dem Messzeitpunkt (links: Mittelwerte; rechts: Mittelwerte, 1 SD, 2 SD; oben: SCL-90-R; mittig: BSS; unten: IIP)

3.5 Soziodemographische Unterschiede der untersuchten Patientengruppen

Um der Frage nachzugehen, ob mögliche Unterschiede der Therapieergebnisse zwischen den untersuchten Gruppen tatsächlich auf die untersuchte Variable Geschlecht oder das Vorliegen einer Persönlichkeitsstörung zurückzuführen sind oder ob soziodemographische Unterschiede konfundierend als Ursache in Betracht kommen, wurde untersucht, ob sich Männer und Frauen bzw. Personen mit und ohne Persönlichkeitsstörungen hinsichtlich soziodemographischer Aspekte unterschieden. Zunächst werden die Ergebnisse für Männer und Frauen dargestellt.

Die Gruppen der männlichen und weiblichen Patienten unterschieden sich signifikant hinsichtlich zahlreicher soziodemographischer Merkmale, beispielsweise dem Familienstand, dem Schul- und Berufsabschluss, der Beschäftigung vor Aufnahme und der Berufstätigkeit, nicht jedoch bezüglich einer Berentung zum Aufnahmezeitpunkt (s. Tabelle 28). Es bestand kein signifikanter Altersunterschied zwischen den Geschlechtern ($t = 1{,}94$, $p = 0{,}052$ n. s.), bei einem Durchschnittsalter bei Aufnahme von 35,9 (SD = 11,6) Jahren der Männer und 35,2 (SD = 11,8) Jahren der Frauen.

Tabelle 28: Soziodemographische Aspekte differenziert nach dem Geschlecht

	Männer		Frauen		Statistik	
	n	%	n	%	χ^2	p
Familienstand					30,832	< 0,001
ledig	982	68,2	1500	61,9		
verheiratet	299	20,8	499	20,6		
geschieden / verwitwet / getrennt lebend	158	11,0	423	17,5		
Schulabschluss					20,776	< 0,001
Hauptschulabschluss	255	17,7	380	15,7		
Realschulabschluss / Mittlere Reife	426	29,7	865	35,8		
Abitur / Fachhochschulreife	674	47,0	1008	41,7		
Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige	78	5,4	165	6,8		
höchster Berufsabschluss					28,858	< 0,001
noch in der Berufsausbildung	114	8,0	242	10,0		
Lehre	465	32,5	796	33,0		
Meister / Fachschule	96	6,7	186	7,7		
Fachhochschule / Universität	276	19,3	430	17,8		
ohne Abschluss	340	23,8	440	18,2		
sonstiger Abschluss	139	9,7	317	13,1		

Tabelle 28 (Fortsetzung): Soziodemographische Aspekte differenziert nach dem Geschlecht

	Männer		Frauen		Statistik	
	n	%	n	%	χ^2	p
Berufstätigkeit vor Aufnahme					122,510	< 0,001
Arbeiter	166	11,6	217	9,0		
Facharbeiter	116	8,1	71	2,9		
einfacher / mittlerer Angestellter / Beamter	255	17,8	624	25,9		
höherer Angestellter / Beamter	132	9,2	133	5,5		
selbständig	73	5,1	72	3,0		
Schüler / Student / Auszubilden-der / Umschüler	302	21,1	502	20,8		
Rentner / Pensionär	117	8,2	284	11,8		
ohne Beruf / Hausfrau / -mann	272	19,0	510	21,1		
Beschäftigung vor Aufnahme					87,325	< 0,001
voll erwerbstätig	523	36,5	725	30,1		
teilzeitbeschäftigt	141	9,9	459	19,1		
nicht erwerbstätig	370	25,9	730	30,4		
arbeitslos	397	27,7	491	20,4		
Berentung					10,376	0,001
ja	153	10,7	345	14,3		
nein	1279	89,3	2069	85,7		
Arbeitsfähigkeit bei Aufnahme					1,699	0,192
arbeits- / schulfähig / teilweise arbeits- / teilweise schulfähig	236	26,0	432	28,4		
arbeits- / schul- / berufs- / er-werbsunfähig	673	74,0	1089	71,6		

Auch für Patienten mit und ohne Persönlichkeitsstörungen wurde untersucht, ob soziodemographische Unterschiede vorlagen. Hierbei zeigte sich, dass die Patienten mit Persönlichkeitsstörungen signifikant jünger (34,3 ± 11,0 Jahre) als die Patienten ohne Persönlichkeitsstörungen (36,6 ± 12,3 Jahre) waren ($t = 5{,}98$, $p < 0{,}001$***). Sie unterschieden sich auch signifikant hinsichtlich des Geschlechts, des Familienstands, des Schul- und Berufsabschlusses, der Berufstätigkeit und der Beschäftigung vor der Aufnahme (s. Tabelle 29). Bei der Betrachtung der soziodemographischen Aspekte ergaben sich deskriptiv die im Folgenden genannten Befunde. Der Anteil an Frauen war niedriger in der Gruppe der Patienten mit Persönlichkeitsstörungen. Die Patienten mit Persönlichkeitsstörungen waren seltener verheiratet, geschieden, verwitwet oder getrennt lebend und häufiger ledig als die Patienten ohne Persönlichkeitsstörungen. Sie wiesen außerdem schlechtere Schul- und Berufsabschlüsse auf. Während die Patienten ohne Persönlichkeitsstörungen häufiger das Abitur oder die Fachhochschulreife erlangt hatten und einen Hochschul- oder Universitätsab-

schluss, eine Meisterprüfung oder einen Fachschulabschluss vorweisen konnten als die Patienten mit Persönlichkeitsstörungen, besaßen diese häufiger Haupt-, Realschulabschlüsse oder die Mittlere Reife und gaben häufiger einen fehlenden Berufsabschluss oder eine Lehre als höchsten Berufsabschluss an. In der Gruppe der Patienten mit Persönlichkeitsstörungen war der Anteil an Arbeitern, Personen ohne Beruf und Hausfrauen/-männern höher und der Anteil an Angestellten und Beamten niedriger als in der Gruppe der Patienten ohne Persönlichkeitsstörungen. Zudem waren die persönlichkeitsgestörten Patienten häufiger arbeitslos oder nicht erwerbstätig als ihre Mitpatienten. Bezüglich einer Berentung oder der Arbeitsfähigkeit bei Aufnahme unterschieden sich die beiden Gruppen nicht signifikant voneinander (s. Tabelle 29).

Tabelle 29: Soziodemographische Aspekte differenziert nach dem Vorliegen einer Persönlichkeitsstörung

	Patienten ohne Persönlichkeitsstörungen		Patienten mit Persönlichkeitsstörungen		Statistik	
	n	%	n	%	χ^2	p
Geschlecht					5,776	0,016
männlich	713	35,5	727	39,2		
weiblich	1296	64,5	1126	60,8		
Familienstand					30,745	< 0,001
ledig	1211	60,3	1271	68,6		
verheiratet	474	23,6	324	17,5		
geschieden / verwitwet / getrennt lebend	323	16,1	258	13,9		
Schulabschluss					17,072	0,001
Hauptschulabschluss	311	15,5	324	17,6		
Realschulabschluss / Mittlere Reife	632	31,5	659	35,7		
Abitur / Fachhochschulreife	939	46,8	743	40,3		
Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige	124	6,2	119	6,4		
höchster Berufsabschluss					47,091	< 0,001
noch in der Berufsausbildung	200	10,0	156	8,5		
Lehre	627	31,4	634	34,4		
Meister / Fachschule	161	8,1	121	6,6		
Fachhochschule / Universität	431	21,6	275	14,9		
ohne Abschluss	351	17,6	429	23,3		
sonstiger Abschluss	229	11,5	227	12,3		

Tabelle 29 (Fortsetzung): Soziodemographische Aspekte differenziert nach dem Vorliegen einer Persönlichkeitsstörung

	Patienten ohne Persönlichkeitsstörungen		Patienten mit Persönlichkeitsstörungen		Statistik	
	n	%	n	%	χ^2	p
Berufstätigkeit vor Aufnahme					83,347	< 0,001
Arbeiter	162	8,1	221	12,0		
Facharbeiter	96	4,8	91	4,9		
einfacher / mittlerer Angestellter / Beamter	521	26,0	358	19,4		
höherer Angestellter, Beamter	174	8,7	91	4,9		
selbständig	83	4,1	62	3,4		
Schüler / Student / Auszubilden-der / Umschüler	438	21,9	366	19,9		
Rentner / Pensionär	192	9,6	209	11,3		
ohne Beruf / Hausfrau / -mann	338	16,9	444	24,1		
Beschäftigung vor Aufnahme					43,005	< 0,001
voll erwerbstätig	724	36,3	524	28,5		
teilzeitbeschäftigt	333	16,7	267	14,5		
nicht erwerbstätig	543	27,2	557	30,3		
arbeitslos	397	19,9	491	26,7		
Berentung					1,682	0,195
ja	246	12,3	252	13,7		
nein	1758	87,7	1590	86,3		
Arbeitsfähigkeit bei Aufnahme					0,090	0,764
arbeits- / schulfähig / teilweise arbeits- / teilweise schulfähig	370	27,3	298	26,7		
arbeits- / schul- / berufs- / er-werbsunfähig	987	72,7	817	73.3		

Aufgrund oben genannter signifikanter soziodemographischer Unterschiede stellte sich die Frage, inwiefern diese die Untersuchung der Therapieergebnisse im Sinne einer Konfundierung störten. Zur Untersuchung, wie stark die soziodemographischen Parameter, hinsichtlich derer sich die Patienten der untersuchten Gruppen unterschieden, die Auftretenswahrscheinlichkeit von Misserfolg oder Erfolg der Therapie beeinflussten, wurde eine logistische Regression durchgeführt. Diese wurde mithilfe der Ergebnisse der SCL-90-R (Erfolg vs. Misserfolg) berechnet. Es gingen 3359 Fälle (87,0 %) in die Analyse ein. Bei den übrigen 503 Fällen (13,0 %) lagen fehlende Werte vor, sodass sie ausgeschlossen wurden. Da der Cut-off zur Zuordnung zu den Gruppen Erfolg oder Misserfolg dem Anteil der Fälle mit der Ausprägung „1“, hier Misserfolg, gleichkommt (Fromm 2005), betrug er in diesem Modell 0,33 (s. Abbildung 5, unter Ausschluss der Pseudogesunden ergab sich $\frac{\text{Verschlechterte + Nonresponder}}{\text{Erfolgreiche+Verschlechterte+Nonresponder}} = \frac{23{,}4\,\%+5{,}9\,\%}{58{,}5\,\%+23{,}4\,\%+5{,}9\,\%} = 0{,}33$). Nur zwischen dem

Alter und dem Familienstand bestand eine hohe Korrelation (r = -0,547), sodass alle Faktoren in die Analyse eingeschlossen werden konnten.

Die Ergebnisse, die das Modell der logistischen Regression beschreiben, sind in Tabelle 30 dargestellt. Der Familienstand, der höchste Schulabschluss, die Beschäftigung vor Aufnahme, das Vorliegen einer Berentung und das Geschlecht wurden als Prädiktoren in das Modell aufgenommen (s. Tabelle 30). Die Variablen Vorliegen einer Persönlichkeitsstörung, Berufstätigkeit vor Aufnahme, höchster Berufsabschluss und Alter bei Aufnahme mit den jeweiligen Referenzkategorien „keine Persönlichkeitsstörung", „ohne Beruf / Hausfrau / -mann" und „Fachhochschule / Universität" führten zu keiner signifikanten Verbesserung in der Vorhersage des Modells und wurden folglich in der logistischen Regresssion, die vorwärts schrittweise erfolgte, ausgeschlossen.

Tabelle 30: Modell der logistischen Regression bezüglich des Einflusses soziodemographischer Aspekte auf das Auftreten von Therapiemisserfolg

		Regressionskoeffizient b	Standardfehler	Wald Statistik	df	Signifikanz	Effektkoeffizient	
							e^b	95 %-Konfidenzintervall
Familienstand	*verheiratet*			6,256	2	0,044		
	ledig	-0,194	0,095	4,147	1	0,042	0,824	0,684-0,993
	geschieden / verwitwet / getrennt	-0,289	0,125	5,33	1	0,021	0,749	0,586-0,957
Höchster Schulabschluss	*Abitur / Fachhochschulreife*			11,497	3	0,009		
	Hauptschulabschluss	0,262	0,109	5,721	1	0,017	1,299	1,048-1,61
	Realschulabschluss / Mittlere Reife	0,27	0,087	9,757	1	0,002	1,31	1,106-1,553
	Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige	0,211	0,156	1,832	1	0,176	1,234	0,91-1,675
Beschäftigung vor Aufnahme	*voll erwerbstätig*			9,907	3	0,019		
	teilzeitbeschäftigt	0,148	0,117	1,583	1	0,208	1,159	0,921-1,459
	nicht erwerbstätig	0,304	0,105	8,446	1	0,004	1,355	1,104-1,663
	arbeitslos	0,246	0,104	5,573	1	0,018	1,279	1,043-1,568
Berentung	ja	0,329	0,118	7,763	1	0,005	1,389	1,102-1,751
Geschlecht	männlich	0,228	0,078	8,558	1	0,003	1,257	1,078-1,464
Konstante		-0,981	0,112	76,339	1	0	0,375	

Die jeweiligen Referenzkategorien sind *kursiv* gedruckt.

Referenzkategorien der dichotomen Variablen: Berentung: *nein*, Geschlecht: *weiblich*

Wie der Tabelle 30 zu entnehmen ist, war beispielsweise die Chance für Misserfolg für Personen mit Haupt- oder Realschulabschluss bzw. mittlerer Reife 1,3-mal so groß wie für Abiturienten (s. e^b unter Berücksichtigung der Signifikanz und des 95 %-Konfidenzintervalls in Tabelle 30). Im Vergleich zum Abitur waren somit ein Haupt- oder Realschulabschluss bzw. die Mittlere Reife Risikofaktoren für Misserfolg. Auch Personen, die arbeitslos oder nicht erwerbstätig waren, im Vergleich zu voll erwerbstätigen Personen, sowie Berentete im Vergleich zu Personen ohne Berentung und Männer im Vergleich zu Frauen hatten eine erhöhte Chance für Misserfolg (im Sinne von Risikofaktoren). Im Vergleich zu den verheirateten Patienten hatten ledige Patienten hingegen eine Chance für Misserfolg von 0,8 und geschiedene, verwitwete oder getrenntlebende Patienten von 0,7 (im Sinne von Schutzfaktoren).

Die Wald-Statistik zeigte, dass mit Ausnahme der Dummy-Variablen „Sonderschule / ohne Schulabschluss / noch in der Schule / sonstige“ und „teilzeitbeschäftigt“ die Regressionskoeffizienten aller aufgenommen Variablen signifikant von Null verschieden waren und diese somit einen signifikanten Einfluss auf das Auftreten von Erfolg bzw. Misserfolg in der Therapie hatten. Der signifikante Omnibus-Test ($\chi^2 = 56{,}7$, $p < 0{,}001$***) zeigte ebenfalls, dass alle in das Modell aufgenommenen Variablen zusammen einen Einfluss auf die abhängige Variable hatten. Die anfängliche -2-Log-Likelihood (Devianz) betrug 4276,6. Sie sank auf 4219,9, was bedeutete, dass die Variablen eine Bedeutung für die Trennung der Gruppen Erfolg vs. Misserfolg haben (Fromm 2005). Nagelkerkes R^2 gibt den „Anteil der Varianz der abhängigen Variable [an], der durch alle unabhängigen Variablen zusammen erklärt wird“ (Fromm 2005: 22). Mit dem berechneten Modell konnten lediglich 2,3 % der Varianz erklärt werden. Dies verwies auf eine schlechte Erklärungskraft des Modells. Durch das Modell wurden insgesamt 54,8 % der Patienten richtig ihren Gruppen zugeordnet. Dabei wurden 52,5 % der erfolgreich therapierten Patienten und 59,3 % der Patienten mit Misserfolg inklusive Nonresponse korrekt erkannt.

4 Diskussion

4.1 Zusammenfassung der Studienergebnisse

Die hier vorgestellte naturalistische Psychotherapiestudie bestimmte auf der Grundlage einer großen Stichprobe und mehrerer Outcome-Parameter die Häufigkeit von Nonresponse und Verschlechterung in der stationären Psychotherapie und untersuchte konkret die Fragen, ob sich Männer und Frauen oder Personen mit und ohne Persönlichkeitsstörungen oder Personen mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen in der Häufigkeit von Misserfolg (Nonresponse und Verschlechterung) unterscheiden.

Je nach Art der Verlaufsmessung (direkt vs. indirekt), der Perspektive (Patienten vs. Therapeuten) und des angewandten Untersuchungsinstrumentes verschlechterten sich 1,9-7,5 % der Patienten und 10,1-50,0 % galten als Nonresponder.

Mithilfe der verwendeten Erfassungsmethoden konnte kein einheitlicher Unterschied in der Auftretenshäufigkeit von Nonresponse und Verschlechterung zwischen Männern und Frauen festgestellt werden. Lediglich der Messung mithilfe der SCL-90-R zufolge wiesen Männer häufiger Misserfolge als Frauen auf.

In der direkten Verlaufsmessung berichteten Patienten mit Persönlichkeitsstörungen häufiger über eine verschlechterte oder unveränderte Symptomatik als Patienten ohne Persönlichkeitsstörungen. In der indirekten Veränderungsmessung unterschieden sich Patienten mit und ohne Persönlichkeitsstörungen nicht im Hinblick auf das Auftreten von Therapiemisserfolg.

Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen unterschieden sich weder in der direkten noch in der indirekten Veränderungsmessung hinsichtlich des Therapieerfolgs voneinander.

Den Befunden der eingesetzten Untersuchungsverfahren zufolge waren Frauen stärker als Männer und Patienten mit Persönlichkeitsstörungen stärker als Patienten ohne Persönlichkeitsstörungen belastet. Patienten mit Cluster-B-Persönlichkeitsstörungen wiesen lediglich einen höheren Gesamtschwereindex auf als Patienten mit Cluster-C-Persönlichkeitsstörungen, während sie sich in den Werten des IIP und des BSS nicht unterschieden.

Alle Verfahren, die zur indirekten Veränderungsmessung verwandt wurden, zeigten Verbesserungen der Werte zwischen dem Aufnahme- und Entlassungszeitpunkt.

4.2 Diskussion der Methoden

In diesem Kapitel werden das Studiendesign, die verwendeten Ein- und Ausschlusskriterien sowie die Erhebungsinstrumente auch im Vergleich zu anderen Psychotherapiestudien beurteilt.

4.2.1 Studiendesign

Es erfolgte eine retrospektive und sekundäre Datenanalyse einer beobachtenden, das heißt nicht-experimentellen und nicht-interventionellen klinischen Studie. Dieses Studiendesign wurde ausgewählt, da sich die Fragestellungen auf natürliche Bedingungen bezogen (Beutel et al. 2010). Solche naturalistischen Studien untersuchen das Psychotherapie-Outcome unter den Bedingungen der klinischen Praxis, also bei realen Patienten mit komplexen Störungsbildern und Therapeuten in wirklichen Therapie-Settings, bei denen keine vorhergehende spezielle Vorbereitung erfolgt, flexible Therapieschemen eingesetzt werden und die Behandlungsdauer nach Maßgabe der klinischen Erfordernisse angepasst wird und nicht starr an standardisierten Vorgängen festgehalten wird (Beutel et al. 2010; Shadish et al. 2000). Dementsprechend weist die hier durchgeführte Studie eine hohe externe Validität bzw. klinische Repräsentativität auf (Beutel et al. 2010). Einer strukturalistischen Sichtweise von Theorien und Hypothesen folgend, nach welcher sich Hypothesen immer auf einen speziellen Kontext beziehen, können naturalistische und randomisiert kontrollierte Studien das gleiche, hohe Evidenzniveau erbringen (Beutel et al. 2010). Bei naturalistischen Studien erfolgt definitionsgemäß keine Randomisierung, weil dadurch der Forschungsgegenstand verfälscht würde (Beutel et al. 2010).

Leichsenring und Rüger (2004) haben Kriterien entwickelt, mit deren Hilfe die Evidenz naturalistischer Studien beurteilt werden kann. Folgende Merkmale, die Leichsenring und Rüger (2004) für Studien mit dem höchsten Evidenzgrad fordern, kann diese Studie erfüllen: (1) das Vorliegen von klaren Ein- und Ausschlusskriterien, (2) eine angemessen große Stichprobengröße für eine gute Teststärke, (3) die Verwendung aktueller diagnostischer Methoden, (4) eindeutig beschriebene statistische Methoden und (5) die Anwendung definierter und klinisch repräsentativer Behandlungen von (6) Patienten mit definierten Krankheitsbildern. Allerdings weist diese Studie kein prospektives Design für eine methodisch hochwertige naturalistische Studie auf. Leichsenring und Rüger (2004) fordern außerdem Maßnahmen zur Sicherung der internen Validität. Dazu zählen sie zusätzliche Designelemente, wie Beobachtungen zu mehreren Zeitpunkten vor der Erhebung oder den Vergleich mit anderen Gruppen. Diese beiden Maßnahmen sind in dieser Arbeit nicht durchgeführt worden. Außerdem gehört Leichsenring und Rüger (2004) zufolge auch die Vorhersage komplexer Ergebnismuster zur Sicherung der internen Validität. Beispielsweise sollte der Einfluss von Interaktionen und Moderatorvariablen auf das Ergebnis berücksichtigt werden (Beutel et al. 2010; Leichsenring und Rüger 2004). Dieser Forderung wird in dem Sinne nachgekommen, als dass die in der Fragestellung untersuchten Gruppen auf

soziodemographische Unterschiede untersucht worden sind. Um den Rahmen dieser Arbeit nicht zu sprengen, konnten jedoch keine weiteren Interaktionen und Moderatorvariablen darüber hinaus untersucht werden. Dies könnte Inhalt weiterer Studien sein. Nichtrandomisierte Vergleichsbedingungen, die Leichsenring und Rüger (2004) zusätzlich als Qualitätsmerkmal von naturalistischen Studien definieren, liegen in dieser Studie nicht vor. Dies ist darauf zurückzuführen, dass im vorhandenen Datensatz, welcher der sekundären Datenanalyse zugrunde liegt, keine Daten zu solchen Vergleichsbedingungen vorliegen. Das wiederum ist durch die schwierige Erzeugung solcher Daten zu begründen. Das Vorenthalten einer Behandlung zur Schaffung einer Vergleichsgruppe wäre beispielsweise ethisch nicht vertretbar. Außerdem wäre die Erhebung von Daten bei Patienten, die aus anderen Gründen keine Therapie erhielten, nach einer Zeitspanne, die der Therapiedauer bei Patienten mit Behandlung entspräche, schwierig zu organisieren und mit großem Aufwand verbunden und damit nicht ökonomisch in den Alltag einer Klinik zu integrieren. Zusätzlich wäre dabei die Vergleichbarkeit der Gruppe der Behandelten und Nicht-Behandelten fragwürdig.

Nach vorausgegangener Diskussion des naturalistischen Studiendesigns folgt an dieser Stelle die kritische Betrachtung weiterer Aspekte des Studienablaufs. Hierzu zählen das untersuchte Therapieverfahren und der Erhebungsort und -zeitpunkt. Im Vergleich dieser Studie mit bestehenden Therapiestudien fällt auf, dass die Ergebnisse teilweise schwer zu vergleichen sind, weil die Beschreibungen des Studienablaufs in den vorliegenden Studien ungenau erfolgen. So sind beispielsweise die angewandten Therapieverfahren häufig unzureichend in den Studien beschrieben. Dadurch wird die konkrete Ausgestaltung des Therapieverfahrens je nach Therapierichtung nicht erkennbar und somit nicht vergleichbar. Außerdem bleibt häufig unklar, ob Unterschiede innerhalb gleicher Ansätze z. B. in der Intensität der Therapie, in der Zusammenstellung der Therapieelemente oder zwischen den Therapieplänen der einzelnen Patienten einer Klinik bestehen (für Studien, in denen psychodynamische Therapieansätze untersucht wurden, vgl. Davies-Osterkamp et al. 1996; Franz et al. 2000; Haase et al. 2008; Konzag et al. 2004; Kriebel et al. 2001; Rudolf et al. 2004; Sammet et al. 2004).

Während die Arbeitsgruppen um Davies-Osterkamp (1996), Franz (2000), Oster (2009) und Nübling (2004) mehrere Erhebungsorte im Sinne einer Multi-Center-Studie genutzt haben, wurden in dieser Studie, wie in den übrigen in Tabelle 31 genannten Studien, lediglich die Daten aus einer Klinik verwendet. Für eine bessere Verallgemeinerbarkeit der Ergebnisse wären Erhebungen an weiteren Krankenhäusern sinnvoll. Dennoch konnten auch im Fachklinikum Tiefenbrunn als alleinigem Erhebungsort ausreichend große Stichproben gewonnen werden (s. 4.2.2). Eine Studie wie die Studie von Dechartres und Kollegen (2011), die für randomisiert kontrollierte Studien der Inneren Medizin gezeigt hat, dass Single-Center-Studien größere Behandlungseffekte messen als Multi-Center-Studien, existiert für naturalistische Studien im Bereich der Psychotherapieforschung nicht. Dadurch ist

nicht bekannt, inwiefern sich die Ergebnisse in einer Multi-Center-Studie unterscheiden würden.

In dieser Arbeit wurden die Ergebnisse zum Entlassungszeitpunkt untersucht. Die Daten erlauben bisher keine Vorhersage, ob das erreichte Therapieergebnis bis zu einem späteren Zeitpunkt stabil bleibt, was zusätzlich von Interesse wäre. Geiser und Mitarbeiter (2001) bewerteten den Therapieerfolg zum Entlassungszeitpunkt lediglich als „mäßig gute[n] Prädiktor [...] für Therapieerfolg bei Katamnese zwei Jahre später" (Geiser et al. 2001: 256). In ihrer Studie verbesserte sich der Großteil der zum Entlassungszeitpunkt der SCL-90-R zufolge verschlechterten Patienten im Katamnesezeitraum hin zu Werten nahe des Werts bei Therapiebeginn (Geiser et al. 2001). Dies könnte auch Ausdruck des Phänomens der Regression zur Mitte bei wiederholter Messung sein (Geiser et al. 2001). Die Aussagekraft der genannten Studie (Geiser et al. 2001) sei jedoch infrage gestellt, da lediglich n = 6 verschlechterte Patienten zum Entlassungszeitpunkt untersucht wurden. Für die Mittelwerte der Summenscores der SCL-90-R und des IIP stellten Junge und Ahrens (1996) hingegen heraus, dass diese in ihrer Studie im zeitlichen Verlauf zwischen dem Entlassungszeitpunkt und dem Katamnesezeitpunkt ein Jahr später stabil geblieben sind. In der Studie von Rudolf und Mitarbeitern (2004) veränderten sich die Therapieergebnisse (gemessen mit dem PSKB-Se-R (Psychischer und Sozialkommunikativer Befund)) unterschiedlich stark zwischen dem Entlassungszeitpunkt und dem Katamnesezeitpunkt sechs bis neun Monate später abhängig von den Diagnosen der Patienten. Die inkonsistenten Aussagen zum zeitlichen Verlauf des Therapieergebnisses, die eine Verbesserung (Geiser et al. 2001), einen stabilen Verlauf (Junge und Ahrens 1996) oder eine Veränderung in Abhängigkeit von der Diagnose (Rudolf et al. 2004) nahelegen, erlauben somit keine allgemeingültige Ableitung einer Prognose für den katamnestischen Verlauf von Therapieergebnissen. Einer der Herausgeber des IIP formulierte, dass größere Effekte erst nach längeren Katamnesezeiträumen im IIP deutlich würden (Strauß 2003), sodass besonders die Ergebnisse des IIP neben denen der anderen Testverfahren zu einem späteren Katamnesezeitpunkt zusätzlich interessant wären.

4.2.2 Ein- und Ausschlusskriterien

In die Studie wurden Patienten eingeschlossen, die vom 01.01.2005 bis zum 31.12.2014 mindestens sechs Wochen im Fachklinikum Tiefenbrunn behandelt wurden, mindestens 18 Jahre alt waren und für die der Global Severity Index der SCL-90-R als Ergebnismaß sowohl bei der Aufnahme als auch bei der Entlassung erfasst wurde. Diese Einschlusskriterien erfüllten 3862 der 6481 in diesem Zeitraum in Tiefenbrunn behandelten Patienten.

Der zehn Jahre umfassende Untersuchungszeitraum vom 01.01.2005 bis zum 31.12.2014 ist mit Ausnahme der Studie von Spitzer und Kollegen (2012), bei der Daten aus elf Jahren verwendet wurden (01.01.2002 bis 31.12.2012), länger und aktueller als der untersuchte Zeitraum anderer Studien zu Misserfolgen in der stationären Psychotherapie. Vergleichs-

weise untersuchten die Arbeitsgruppen um Franz (2000) Daten aus acht Jahren zwischen April 1990 und Juli 1998, um Geiser (2001) einen Zeitraum von zwei Jahren, um Haase (2008) drei Jahre vom ersten Quartal 2001 bis vierten Quartal 2003, Junge und Ahrens (1996) ein Jahr von 1991 bis1992, Konzag und Kollegen (2004) die Jahre 1995 bis 2000, Kriebel und Mitarbeiter (2001) Patienten aus dem Jahr 1996, Oster und Mitautoren (2009) knapp ein Jahr von Juni 2006 bis April 2007, Rudolf und Kollegen (2004) fünf Jahre von 1996 bis 2000 und Sammet und Mitarbeiter (2004) Daten von 1998 bis 2002. Die Beobachtung über solch einen langen Zeitraum ermöglichte das Erfassen einer großen Stichprobe. Mit 3862 Patienten war die Untersuchungsstichprobe größer als in anderen Studien zu Misserfolgen in der stationären Psychotherapie (s. Tabelle 31). Selbst bei der Messung mit dem BSS, für den in dieser Studie die niedrigste Anzahl an vollständig dokumentierten Fällen bestand, lagen Ergebnisse von 2148 Patienten vor. Eine solch hohe Probandenzahl weist keine der Studien zu Misserfolgen in der stationären Psychotherapie auf (s. Tabelle 31).

Der Vorteil einer großen Stichprobe wie dieser ist, dass selbst bei der Betrachtung mehrerer Untergruppen diese für statistische Verfahren wie z. B. den χ^2-Test noch ausreichend groß sind. Somit war die Voraussetzung, dass erwartete Häufigkeiten kleiner als 5 höchstens in einem Fünftel der Zellen einer Tabelle vorliegen (Bühl 2016; Kuckartz et al. 2013), stets erfüllt. Das zentrale Grenzwerttheorem besagt, dass sich die Mittelwerte von Stichproben mit einer Größe von $n \geq 30$ um den Mittelwert der Grundgesamtheit normalverteilen, unabhängig von der Verteilung der Rohwerte (Bortz und Schuster 2010; Döring und Bortz 2016b; Kuckartz et al. 2013). Bei der Beurteilung mithilfe der SCL-90-R (mithilfe des IIP) [mithilfe des BSS] betrug die Stichprobengröße der Männer n = 1440, (n = 1418), [n = 799], der Frauen n = 2422, (n = 2396), [n = 1349], der Gruppe an Patienten mit Persönlichkeitsstörungen n = 1853, (n = 1820), [n = 969] und der Patientengruppe ohne Persönlichkeitsstörungen n = 2009, (n = 1994), [n = 1179], der Gruppe an Patienten mit Cluster-B-Persönlichkeitsstörungen n = 896, (n = 879), [n = 451] und der Gruppe an Patienten mit Cluster-C-Persönlichkeitsstörungen n = 884, (n = 868), [n = 470]. Damit konnte bei der Untersuchung des Zusammenhangs zwischen der Gruppenzugehörigkeit und den Summenscores der indirekten Veränderungsmessung die Voraussetzung der Normalverteilung als erfüllt angenommen werden. Die großen Stichproben waren auch bei der Anwendung der t-Tests beim Stichprobenvergleich oder im Rahmen der direkten Veränderungsmessung (Therapeutenperspektive: Frauen n = 1387, Männer n = 817, Patienten mit Persönlichkeitsstörungen n = 990, Patienten ohne Persönlichkeitsstörungen n = 1214, Patienten mit Cluster-B-Persönlichkeitsstörung n = 465, Patienten mit Cluster-C-Persönlichkeitsstörung n = 477; Patientenperspektive: Frauen n = 2377, Männer n = 1427, Patienten mit Persönlichkeitsstörungen n = 1817, Patienten ohne Persönlichkeitsstörungen n = 1987, Patienten mit Cluster-B-Persönlichkeitsstörung n = 871, Patienten mit Cluster-C-Persönlichkeitsstörung n = 873) vorteilhaft, da t-Tests bei Stichprobengrößen von

$n \geq 30$ ihr festgelegtes Signifikanzniveau auch bei nicht normalverteilten Merkmalen einhalten (Bortz und Schuster 2010).

Stärke und Schwäche zugleich ist, dass bei größeren Stichproben die Teststärke steigt, sodass mit größerer Wahrscheinlichkeit auch kleine Effekte erkannt werden (Bortz und Schuster 2010). Dadurch muss jedoch davon ausgegangen werden, dass solche Nullhypothesen wie in dieser Arbeit, die besagen, dass die Mittelwerte zweier Gruppen identisch sind, bei ausreichend großen Stichproben verworfen werden müssen (Bortz und Schuster 2010). Es besteht also die Möglichkeit, dass sich ein signifikanter Unterschied darstellen lässt, der jedoch keine praktische Bedeutung hat. Wird das Ergebnis eines Signifikanztests jedoch unter Berücksichtigung der praktischen Bedeutsamkeit interpretiert (Bortz und Schuster 2010), wie es in dieser Arbeit der Fall ist, überwiegt der Vorteil einer großen Stichprobe.

Es wurden nur Patienten mit einer Mindestbehandlungsdauer von 6 Wochen in die Studie eingeschlossen. Dies lässt sich einerseits damit begründen, dass eine Kurzzeittherapie im Fachklinikum Tiefenbrunn mit dem Erreichen des 42. Tages als vollständig gilt, und damit, dass es sinnvoll ist, ein Behandlungskonzept erst zu einem Zeitpunkt zu evaluieren, wenn Veränderungen bereits eingetreten und auf die Therapie zurückführbar sind. Dieselbe Mindestbehandlungsdauer von 42 Tagen verwendeten Spitzer und Kollegen (2012) in ihrer Studie, wobei dort zusätzlich eine reguläre Beendigung der Therapie vonseiten des Therapeuten und des Patienten als Einschlusskriterium galt. Auch Junge und Ahrens (1996) forderten eine regelhafte Beendigung als Kriterium. Die Arbeitsgruppe um Franz (2000) wählte zehn Tage, um Frühabbrecher und Kurzlieger auszuschließen. Geiser und Mitarbeiter (2001) definierten vier Wochen und Sammet und Mitarbeiter (2004) acht Wochen als Einschlusskriterium ohne ihre Wahl zu begründen. Da Therapieabbrüche bzw. die vorzeitige Beendigung der Behandlung in der Literatur als eine Form von Misserfolg in der Psychotherapie diskutiert werden (Jacobi 2001; Reuter und Scheidt 2014), wurde die reguläre Beendigung der Therapie in dieser Arbeit nicht als Einschlusskriterium festgelegt, damit die Häufigkeiten für Nonresponse und Verschlechterung als eine Variante von Misserfolg auch für Patienten miterfasst wurden, die mit einem Therapieabbruch eine andere Form von Misserfolg aufwiesen. Da jedoch nur 0,7 % der Patienten der Stichprobe die Behandlung gegen ärztlichen Rat beendeten, fallen ihre Ergebnisse kaum ins Gewicht. In einer Fall-Kontroll-Studie von Martens und Mitarbeitern (2014), die 59 Patienten, die die Therapie gegen ärztlichen Rat innerhalb von 28 Tagen abbrachen, und 59 Kontrollpatienten untersuchten, unterschieden sich die Abbrecher nicht signifikant von der Kontrollgruppe hinsichtlich der Ergebnisse der SCL-90-R und des IIP. Somit waren keine bedeutsamen Unterschiede zwischen Patienten mit regulärer Therapiebeendigung und Patienten, die die Therapie gegen ärztlichen Rat beendeten, zu erwarten. Dies rechtfertigt, dass der vergleichsweise hohe Anteil an Abbrechern (358 von 1361 Patienten (26,3 %)) in der Patientengruppe, die aufgrund einer Therapiedauer von unter 42 Tagen aus der Studie ausgeschlossenen wurden, in dieser Arbeit nicht weiter untersucht wurde.

Das Einschlusskriterium, dass der Global Severity Index der SCL-90-R sowohl bei der Aufnahme als auch bei der Entlassung erfasst sein musste, wurde formuliert, damit für alle untersuchten Patienten mindestens ein Ergebnismaß zur Beurteilung des Therapieerfolgs vorlag. Da der GSI häufiger als der Gesamtwert des BSS oder des IIP und auch häufiger als die direkte Messung vollständig erfasst wurde, wurden vollständige GSI-Werte als Einschlusskriterium gewählt. Außerdem wurde die SCL-90-R in anderen Studien besonders häufig als Fragebogen verwendet (s. 4.2.3), was eine Vergleichbarkeit mit diesen ermöglicht. Wünschenswert wären selbstverständlich auch vollständige Werte für die anderen Erhebungsarten.

Das Mindestalter von 18 Jahren wurde als Einschlusskriterium gewählt, um eine Vergleichbarkeit mit anderen Psychotherapie-Studien zu gewährleisten.

4.2.3 Untersuchungsinstrumente

Alle in dieser Arbeit im Rahmen der indirekten Veränderungsmessung verwendeten Untersuchungsinstrumente (SCL-90-R, IIP, BSS) gelten als objektive, reliable und gut validierte Tests und sind explizit für eine Verlaufsmessung zugelassen (s. 2.3). Stieglitz und Freyberger (2001) kritisieren, dass Selbstbeurteilungsverfahren häufig bezüglich des Zeitpunktes oder der Fragestellung falsch verwendet werden. Da beispielsweise im Manual zur SCL-90-R jedoch explizit eine Verwendung zur Verlaufsbeurteilung als Prä-Post-Messung vorgesehen ist (Franke 2002), wie sie in dieser Arbeit durchgeführt wurde, gilt dieser Vorwurf nicht für diese Erhebung.

In der Literatur besteht teilweise die Forderung, störungsspezifische Beurteilungsverfahren zu verwenden, da störungsübergreifende Tests spezielle Bereiche einer Störung nicht ausreichend messen könnten und ihre Validität für die spezifischen Störungen nicht bewiesen sei (Stieglitz und Freyberger 2001). In dieser Studie wurden dennoch bewusst störungsübergreifende Tests eingesetzt, da die Patienten, welche eine stationäre Psychotherapie erhalten, meist mehr als eine Diagnose aufweisen (s.3.1.2).

Durch die Kombination von Selbst- und Fremdbeurteilungsverfahren in dieser Arbeit können einerseits die in Kapitel 1.3.4 genannten Vorteile beider Verfahren kombiniert werden und andererseits fallen die für die Selbst- und Fremdbeurteilung spezifischen Fehler geringer ins Gewicht, da sie in der jeweils anderen Erhebungsart nicht auftreten. Zu solchen spezifischen Fehlern für die Selbstbeurteilungsverfahren gehören unbewusste Fehler durch Selbsttäuschung und absichtliche Verfälschungen durch Dissimulation, Simulation und Bagatellisierung, sowie die Tendenz, entsprechend der sozialen Erwünschtheit zu antworten (Möller 1990; Stieglitz 1994; Stieglitz und Freyberger 2001). Spezifisch bei Fremdbeurteilungsverfahren auftretende Fehler können durch falsche Schlussfolgerungen entstehen. Als Beispiele wären logische Fehler oder der Halo-Effekt, bei welchem die Beurteilung einzelner Items durch den Gesamteindruck beeinflusst wird, zu nennen. Auch könnte die Erwartung einer erfolgreichen Therapie durch einen Therapeuten eine Unter-

schätzung von Symptomen im Zeitverlauf bedingen (Möller 1990; Stieglitz 1994; Stieglitz und Freyberger 2001).

Neben der Kombination von Selbst- und Fremdbeurteilungsverfahren wurden in dieser Arbeit jeweils Instrumente der direkten und indirekten Veränderungsmessung angewendet, um der Forderung nach einer multimodalen Diagnostik (s. 1.4) umfassend nachzukommen. Die Vorteile und Einschränkungen der direkten Veränderungsmessung sind in Kapitel 1.3.1 dargestellt. Einige der dort genannten Nachteile entfallen in dieser Arbeit jedoch. Der Nachteil der direkten Veränderungsmessung, dass ohne zusätzliche Informationen nicht bekannt ist, von welchem Niveau aus die Veränderungen beginnen (Baumann et al. 1980; Grawe 1982), liegt beispielsweise nicht vor, da durch die globalen Kennwerte der zusätzlichen indirekten Veränderungsmessung das Ausmaß der Erkrankung bekannt ist. Auch treten die befürchteten Auswertungsprobleme bei Mehrpunkterhebungen (Baumann 1982; Grawe 1982) nicht auf, da nur ein Erhebungszeitpunkt vorliegt. Mit mindestens 42 Tagen Behandlungsdauer ist das Zeitintervall lang genug, sodass auch die Einschränkung, dass die direkte Veränderungsmessung bei kurzen Zeitintervallen weniger geeignet sei (Baumann 1982), nicht relevant ist.

Die Verwendung der SCL-90-R, einem Instrument, das bereits mehrfach in Studien zur Gruppeneinteilung in Verbesserte, Nonresponder und Verschlechterte zur Beurteilung des Therapieerfolgs genutzt wurde, ermöglicht einen Vergleich der Ergebnisse dieser Studie mit denen anderer Studien (s. Tabelle 35). Durch die zusätzliche Verwendung des IIP, des BSS und der direkten Veränderungsmessung übersteigt die Erhebung den Umfang bisheriger Studien und ist die erste Arbeit, die Misserfolge in der Psychotherapie auf Einzelfallebene unter Berücksichtigung sowohl der Selbst- und Fremdbeurteilung als auch der direkten und indirekten Messung und zusätzlich unter Beachtung interpersoneller Aspekte untersucht.

Im Rahmen der indirekten Veränderungsmessung dieser Arbeit wurden lediglich die Summenscores der SCL-90-R, des IIP und des BSS ausgewertet. Auf die zusätzliche Auswertung der einzelnen Skalen oder der einzelnen Items musste leider verzichtet werden, da es den Umfang dieser Arbeit überstiegen hätte. Eine solche Auswertung könnte jedoch gegebenenfalls im Rahmen weiterer Studien interessante Ergebnisse hervorbringen. Dass hier die Summenscores untersucht wurden, begründet sich inhaltlich darin, dass das Auftreten von Nonresponse und Verschlechterung in der stationären Psychotherapie zunächst allgemein untersucht werden sollte. Für die Verwendung der Summenscores sprachen zusätzlich ihre guten Testgütekriterien (s. 2.3.1-2.3.3). So empfiehlt Franke (2002) beispielsweise explizit aufgrund der hohen Reliabilität der Gesamtskala zuerst die Veränderung des GSI auszuwerten. Auch die besonders hohe Reliabilität des BSS-Gesamtwerts (Koo und Li 2016; Schellberg 1995) (s. 2.3.3) spricht für seine Verwendung.

Die Arbeitsgruppe um Stieglitz (2001) weist darauf hin, dass eine Verwendung jahrzehntealter Instrumente, die im Verlauf nicht revidiert worden sind, kritisch zu beurteilen ist, da sie keine neuen Ergebnisse berücksichtigen und ggf. veraltete Normen nutzen würden.

Dies ist bei den hier verwendeten Tests unproblematisch, da jeweils die zum Zeitpunkt der Erhebung aktuellste Version eingesetzt worden ist. Der zu Beginn der Erhebung älteste, jedoch nicht veraltete Test ist mit 10 Jahren der BSS (Schepank 1995), während die Versionen der SCL-90-R (Franke 2002) und des IIP (Horowitz et al. 2000) deutlich aktueller sind.

4.3 Diskussion der Ergebnisse

In diesem Kapitel werden die Repräsentativität der Stichprobe sowie die Ergebnisse der einzelnen Forschungsfragen diskutiert.

4.3.1 Repräsentativität der Stichprobe

Um die Repräsentativität der Stichprobe zu beurteilen, werden im Folgenden soziodemographische und klinische Charakteristika dieser Studie im Vergleich zu bereits veröffentlichten Psychotherapiestudien erörtert. Außerdem werden die Ergebnisse des Vergleichs zwischen der untersuchten Stichprobe und der Gesamtstichprobe vor Anwendung der Einschlusskriterien diskutiert.

4.3.1.1 Soziodemographische Aspekte

Bei der Betrachtung der Zusammensetzung der Stichprobe, fällt der höhere Frauenanteil mit 62,7 % der Patienten im Vergleich zu einem weiblichen Anteil an der Gesamtbevölkerung in Deutschland von 50,9 % (Statistisches Bundesamt (Destatis) 2016b) auf. Der Frauenanteil ist jedoch ähnlich hoch wie in anderen Studien zur stationären Psychotherapie, wo er zwischen 55,2 % (Liebler et al. 2004) und 72,1 % (Geiser et al. 2001) schwankt (s. Tabelle 31).

Tabelle 31: Stichprobengröße, Frauenanteil und Altersdurchschnitt anderer Studien zu Misserfolgen in der stationären Psychotherapie

Studie	Stichprobengröße (n)	Frauenanteil (%)	Alter (Jahre)	
			M	SD
Bassler 1995	218	63,3	34,3	9,7
Davies-Osterkamp et al. 1996	194	61,3	-	-
Franz et al. 2000	495	61,6	37,7	11,5
Geiser et al. 2001	43	72,1	32	9,8
Geyer et al. 2003	61	59	-	-
Haase et al. 2008	156	68,6	-	-
Junge und Ahrens 1996	164	68,9	39,7	11,5
Konzag et al. 2004	254	71	32	9,6
Kriebel et al. 2001	741	63,1	41,7	10,4
Liebler et al. 2004	87	55,2	38,2	11,8
Nübling et al. 2004	858	56,8	43,8	10,6
Oster et al. 2009	463	59	47,3	9,1
Rief et al. 1994	1377	70	-	-
Rudolf et al. 2004	461	75	30	-
Sammet et al. 2004	213	70,9	33,4	12,2
Schauenburg et al. 2001	71	65	35,3	11,7
Spitzer et al. 2012	1239	60	39,7	11,7
Diese Studie	3862	62,7	35,5	11,8

- : fehlende Angaben

Mit einem Durchschnittsalter von 35,5 (SD = 11,8) Jahren kann das Alter der Patienten in der Stichprobe im Vergleich zu den in Tabelle 31 genannten Studien ebenfalls als repräsentativ angesehen werden. Dort lag das Durchschnittsalter minimal bei 30 Jahren (Rudolf et al. 2004) und maximal bei 47,3 Jahren (Oster et al. 2009).

4.3.1.2 Klinische Merkmale

In diesem Kapitel werden die Behandlungsdauer unter Berücksichtigung des Therapiekonzepts, die zwei häufigsten Diagnosen, Vorbehandlungen und Suizidalität als klinische Charakteristika der Stichprobe kritisch betrachtet. Die mittlere Behandlungsdauer der Patienten dieser Studie von 89,7 Tagen (SD = 30,1 Tage) ist vergleichbar mit der Behandlungsdauer in anderen stationären Psychotherapiestudien (s. Tabelle 32). Zwischen der in den anderen Studien angegebenen Behandlungsdauer und dem angewandten Therapiekonzept (s. Tabelle 32) ist kein eindeutiger Zusammenhang erkennbar, was sich dadurch erklären lassen

könnte, dass sich die einzelnen Verfahren der psychodynamischen Psychotherapien nicht strikt voneinander abgrenzen lassen, sondern sich die Übergänge fließend gestalten (Rudolf und Rüger 2001) (s. 1.1). Die Studien, die deutlich niedrigere durchschnittliche Verweildauern (unter 50 Tage) aufweisen (Kriebel et al. 2001; Nübling et al. 2004; Oster et al. 2009), beziehen sich auf die Ergebnisse aus der psychosomatischen Rehabilitation, sodass der Unterschied auf die unterschiedlichen Institutionen zurückführbar sein kann. Außerdem erhöhte sich in dieser Studie die mittlere Verweildauer durch das Einschlusskriterium, dass nur Patienten mit einer Behandlungszeit von mehr als 42 Tagen eingeschlossen wurden, im Vergleich zur Stichprobe vor Anwendung der Ein- und Ausschlusskriterien (M = 72,9 Tage, SD = 38,3 Tage). Dass die mittlere Behandlungsdauer in der Studie von Spitzer und Mitarbeitern (2012) länger war als im Rahmen dieser Arbeit, obwohl die Daten in derselben Klinik erhoben wurden und die Mindestbehandlungsdauer von 42 Tagen ebenfalls als Einschlusskriterium formuliert wurde, könnte dadurch bedingt sein, dass dort nur Patienten mit Persönlichkeitsstörungen in die Studie eingeschlossen wurden. Diese Patienten wiesen auch in dieser Studie mit 94,1 (SD = 28,5) Tagen eine längere mittlere Behandlungsdauer als jene Patienten ohne Persönlichkeitsstörungen mit 85,6 (SD = 31,0) Tagen auf.

Tabelle 32: Behandlungsdauer und Therapiekonzept anderer Psychotherapiestudien

Studie	Behandlungsdauer (Tage)			Therapiekonzept
	M	SD	Bereich	
Bassler 1995	-	-		psychoanalytisch orientiert
Davies-Osterkamp et al. 1996	117	-	14-284	psychoanalytisch orientiert
Franz et al. 2000	126,6	90,2	-	tiefenpsychologisch-psychodynamisch
Geiser et al. 2001	73,5	30,1	-	psychoanalytisch orientiert
Geyer et al. 2003	-	-	-	tiefenpsychologisch fundiert
Haase et al. 2008	60,5	24,5	8-128	psychodynamisch
Junge und Ahrens 1996	66,8	20,6	-	tiefenpsychologisch orientiert, vorwiegend einzeltherapeutisch
Konzag et al. 2004	72	-	-	psychodynamisch
Kriebel et al. 2001	47,9	12,7	-	psychodynamisch
Liebler et al. 2004	93,82	39,57	-	tiefenpsychologisch orientiert
Nübling et al. 2004	49,6	17,7	-	psychosomatische Rehabilitation ohne Angabe des Therapie-Konzepts
Oster et al. 2009	33,3	7,8	-	psychosomatische Rehabilitation ohne Angabe des Therapie-Konzepts
Rief et al. 1994	-	-	-	lernpsychologisch und problem-orientiert
Rudolf et al. 2004	76	-	-	psychodynamisch
Sammet et al. 2004	-	16,31	56-147	psychodynamisch
Schauenburg et al. 2001	-	-	-	psychodynamisch
Spitzer et al. 2012			-	psychodynamisch
Erfolgsgruppe	104,2	27,7		
Misserfolgsgruppe	100,9	23		
Diese Studie	89,7	30,1	42-252	psychodynamisch orientiert

- : fehlende Angaben

Die Verteilung der Diagnosen ist schlecht mit denen anderer Studien vergleichbar. Dies ist darauf zurückzuführen, dass sich beispielsweise die Häufigkeitsangaben in den Studien auf unterschiedlich genau differenzierte Diagnosen beziehen, unterschiedliche Diagnosen zu verschiedenen Gruppen zusammengefasst worden sind oder teilweise nur der Anteil an den Hauptdiagnosen angegeben ist. In dieser Arbeit wurde jedoch der Anteil an Patienten, die eine bestimmte Diagnose aufweisen, unabhängig davon, ob es eine Haupt- oder Nebendiagnose war, bestimmt. Beispielsweise wurden affektive Störungen (F30-39) bei 84,0 % der Patienten als häufigste Diagnosegruppe dokumentiert. Die häufigste genauer differenzierte, zweistellige ICD-Diagnose war die Diagnose F33 (rezidivierende depressive Störungen), die 70,4 % der Patienten aufwiesen. In anderen Studien wurde die Häufigkeit von Diagnosen

dieser genannten Diagnosegruppe in der in Tabelle 33 aufgezeigten Weise angegeben, sodass eine direkte Vergleichbarkeit nicht gegeben ist.

Tabelle 33: Relative Häufigkeiten von Patienten mit Persönlichkeitsstörungen und Diagnosen im Bereich der affektiven Störungen in anderen Studien

Studie	Anteil an Patienten mit Persönlichkeitsstörungen (%)	Anteil an Patienten mit Diagnosen im Bereich der affektiven Störungen (%)
Bassler 1995	11,2 (Hauptdiagnose)	20,2 (Hauptdiagnose Dysthymia)
Davies-Osterkamp et al. 1996	15,6 (Hauptdiagnose)	-
Franz et al. 2000	-	23,1 (Hauptdiagnose Dysthymia) 10,4 (Hauptdiagnose längere Depression)
Geiser et al. 2001	-	23,3 (depressive Störung als Hauptdiagnose, weitere als Komorbidität)
Geyer et al. 2003	-	14,6 (depressive Episode; fehlende Angabe, ob Hauptdiagnose oder Komorbidität)
Haase et al. 2008	-	32 (Hauptdiagnose F32-F34)
Junge und Ahrens 1996	-	-
Konzag et al. 2004	100	33,9 (Komorbidität affektive Störung)
Kriebel et al. 2001	-	-
Liebler et al. 2004	51,72	45,98 (depressive Syndrome: F32,33,31.3 als Haupt- oder Nebendiagnose)
Nübling et al. 2004	3,2 (Hauptdiagnose) 7,6 (Komorbidität)	35,9 (Hauptdiagnose depressive Syndrome: F32,F33,F34.1)
Oster et al. 2009	-	32 (Hauptdiagnose aus dem depressiven Formenkreis)
Rief et al. 1994	-	-
Rudolf et al. 2004	45	37 (Anteil affektiver Störungen an den vier häufigsten Diagnosegruppen)
Sammet et al. 2004	5,6 (Hauptdiagnose) 46 (Komorbidität)	51,11 (Hauptdiagnose depressive Episode: F32, F33)
Schauenburg et al. 2001	49,2	54,9 (Hauptdiagnose depressive Störung)
Spitzer et al. 2012	100	85,3 (Komorbidität affektive Erkrankungen)

- : fehlende Angaben

Lediglich in den beiden Studien, die nur Patienten mit Persönlichkeitsstörungen untersuchten (Konzag et al. 2004; Spitzer et al. 2012), und in der Studie von Liebler und Mitarbeitern

(2004) wurden auch die Komorbiditäten berücksichtigt. Die ähnliche Verteilung der Diagnosen in dieser Arbeit und in der Studie von Spitzer und Mitarbeitern (2012) lässt sich dadurch erklären, dass sich Teile der untersuchten Stichproben überschneiden. In der Studie von Liebler und Mitarbeitern (2004) waren mit durchschnittlich 1,94 Diagnosen pro Patient weniger Diagnosen dokumentiert als in dieser Arbeit (M = 3,03 Diagnosen/Patient). Dies könnte erklären, warum der dort angegebene Anteil an Patienten mit depressiven Syndromen (F32, F33, F31.3), der bei 46,0 % der Patienten lag (Liebler et al. 2004), niedriger war als in dieser Studie, in der 70,4 % der Patienten die Diagnose F 33 aufwiesen. Es bleibt somit offen, ob der Anteil an Patienten mit rezidivierenden depressiven Störungen im Fachklinikum Tiefenbrunn höher war als in anderen Therapieeinrichtungen. Die Arbeitsgruppe um Konzag (2004) nennt die Anzahl der dokumentierten Diagnosen nicht, sodass ein Vergleich nicht möglich ist. Der Anteil an affektiven Störungen als Komorbidität ist mit 33,9 % (Konzag et al. 2004) jedoch niedrig, bzw. der Anteil an affektiven Störungen im Fachklinikum Tiefenbrunn (84,0 %) vergleichsweise hoch.

Die mit 48,5% zweithäufigste Diagnose der Stichprobe (F60, spezifische Persönlichkeitsstörung), die für die dritte Forschungsfrage relevant ist, wurde ähnlich häufig in den Studien der Arbeitsgruppen um Rudolf (2004), Sammet (2004), Schauenburg (2001) und Liebler (2004) genannt (s. Tabelle 33). Dass die Anteile in der Studie von Bassler (1995) und Davies-Osterkamp und Kollegen (1996) deutlich niedriger ausfielen, ist darauf zurückzuführen, dass nur der Anteil an Hauptdiagnosen dokumentiert ist und nicht der Anteil an Patienten, die eine Persönlichkeitsstörung als Haupt- oder Nebendiagnose aufwiesen.

Bei der Interpretation der Daten zur Vorbehandlung der Patienten muss zunächst kritisch angemerkt werden, dass diese nicht für alle Patienten erfasst wurden (s. Tabelle 10). Der wahre Anteil an Patienten, die eine Vorbehandlung erhalten haben, wird durch die dokumentierten Werte überschätzt, da einige Therapeuten vermutlich bei den Fragen nach Vorbehandlungen das entsprechende Feld frei gelassen haben, anstatt aktiv „nein“ zu dokumentieren. Dies kann rückblickend nicht von den fehlenden Angaben unterschieden werden. Kriebel und Kollegen (2001) und Rief und Mitarbeiter (1994) berichten deutlich niedrigere Werte (Kriebel et al. 2001: 58,2 % ambulante und 37,3 % stationäre psychotherapeutische Vorbehandlung. Rief et al. 1994: 53 % keine Vorbehandlung, 29 % eine ambulante psychotherapeutische Vorbehandlung, 19 % mindestens zwei psychotherapeutische Vorbehandlungen in der Vorgeschichte und 38 % stationäre Vorbehandlungen). Die Daten von Bassler (1995) sind mit nur 1,8 % fehlenden Angaben für psychotherapeutische Vorbehandlungen und 1,4 % fehlenden Werten für psychiatrische Vorbehandlungen vollständiger als die Daten aus dem Fachklinikum Tiefenbrunn. Bassler (1995) gab an, dass insgesamt 55,5 % der Patienten psychiatrisch oder psychotherapeutisch vorbehandelt waren.

Ohne Berücksichtigung fehlender Werte waren für 63,1 % der 3862 Patienten eine psychotherapeutische oder psychiatrische oder beide Formen der Vorbehandlung dokumentiert. Obwohl der so berechnete Anteil den wahren Anteil noch unterschätzt, war er höher als in

den Studien von Bassler (1995) und Rief und Kollegen (1994). Dies könnte so interpretiert werden, dass im Fachklinikum Tiefenbrunn im genannten Untersuchungszeitraum schwerer erkrankte Patienten und/oder jene, bei denen eine Therapie in anderen Kliniken zuvor nicht erfolgreich war, behandelt worden sind. Aufgrund fehlenden Vergleichsmaterials kann dies jedoch nicht verifiziert werden. Für eine höhere Krankheitsschwere der Behandelten würde der mit 47,6 % hohe Anteil an Patienten, die im Vorfeld der Therapie suizidal waren, sprechen. Jedoch fehlen auch hier Vergleichswerte anderer Psychotherapiestudien, da lediglich die Arbeitsgruppe um Spitzer (2012) einen niedrigeren Anteil an Patienten nennt, die im Vorfeld der Therapie suizidal waren (36,2 % (Spitzer et al. 2012)). Kritisch anzumerken bleibt, dass fraglich ist, wie präzise die Patienten selber angeben können, welche Art von Therapie sie zuvor erhalten haben.

4.3.1.3 Vergleich der Stichproben vor und nach Anwendung der Einschlusskriterien

Die Anwendung von Ein- und Ausschlusskriterien kann die Repräsentativität von Ergebnissen naturalistischer Studien einschränken (Spitzer et al. 2012). Dass sich die Teilstichprobe hinsichtlich soziodemographischer Aspekte mit Ausnahme des höchsten Schulabschlusses und hinsichtlich des Vorliegens einer Persönlichkeitsstörung nicht von der Stichprobe vor Anwendung der Ein- und Ausschlusskriterien unterschied (s. 3.1.3), spricht jedoch für die Repräsentativität der Stichprobe nach Anwendung der Einschlusskriterien und dafür, dass die Ein- bzw. Ausschlusskriterien die Ergebnisse nicht verzerren.

Der Anteil an Patienten mit höherem Schulabschluss, z. B. Abitur oder Fachhochschulreife, war zwar in der Teilstichprobe höher als in der Gesamtstichprobe, die Beträge der standardisierten Residuen waren jedoch kleiner als 2 und gelten damit nicht als signifikant und relevant (Bühl 2016). Das mit einem Betrag von 1,6 höchste standardisierte Residuum lag für das Abitur / die Fachhochschulreife in der Teilstichprobe vor. Hinzu kommt, dass von dem unterschiedlichen Anteil an Probanden mit Abitur kein Einfluss auf das Therapieergebnis zu erwarten war, wenn man der Studie von Bassler (1995), in der das Abitur nicht als prognostischer Faktor herangezogen werden konnte, oder den Studien von Kriebel und Kollegen (2001) sowie Spitzer und Mitarbeitern (2012), die keinen Zusammenhang zwischen dem Schulabschluss und dem Therapieergebnis fanden, folgt. Dass sich in der Teilstichprobe dieser Arbeit mehr Abiturienten als in der Gesamtstichprobe befanden, lässt sich teilweise durch das Einschlusskriterium der Mindestbehandlungsdauer von 6 Wochen erklären. Bei den Patienten mit einer Behandlungsdauer von unter 42 Tagen, welche also aus der Studie ausgeschlossen wurden, war der Anteil an Patienten mit Abitur signifikant niedriger als bei statistischer Unabhängigkeit zu erwarten, sodass der Anteil an Abiturienten, welche in die Studie eingeschlossen wurden, entsprechend hoch war (χ^2 = 63,9, $p < 0{,}001$ ***). Hierzu passende Beobachtungen machten Franz und Kollegen (2000): Patienten mit (Fach-)Abitur wiesen in ihrer Studie eine längere Aufenthaltsdauer auf als Patienten mit Haupt- oder Realschulabschluss.

Bei der Betrachtung der Patientengruppen, eingeteilt danach, ob eine Cluster-B- oder eine Cluster-C-Persönlichkeitsstörung vorlag, fällt auf, dass Patienten mit Cluster-C-Persönlichkeitsstörung in dieser Arbeit signifikant häufiger in der Stichprobe nach Anwendung der Einschlusskriterien waren als erwartet (χ^2 = 13,408, p = 0,001**). Das lässt sich vorrangig darauf zurückführen, dass signifikant weniger Patienten mit Cluster-C-Persönlichkeitsstörung aufgrund einer Behandlungsdauer von unter 42 Tagen ausgeschlossen wurden, als bei statistischer Unabhängigkeit erwartet worden wäre (χ^2 = 103,569, p < 0,001***, standardisiertes Residuum -7,8). Auch wurden signifikant weniger Patienten mit Cluster-C-Persönlichkeitsstörung aufgrund unvollständiger SCL-90-R-Werte ausgeschlossen (χ^2 = 43,565, p < 0,001***, standardisiertes Residuum -3,9), wobei erstgenannter Aspekt stärker zur Erhöhung der Anzahl der Patienten mit Cluster C-Persönlichkeitsstörung in der Stichprobe nach Anwendung der Einschlusskriterien beitrug (s. standardisierte Residuen). Dieses Ergebnis ist vor dem Hintergrund der Charaktereigenschaften, die in der Gruppe der Cluster-C-Persönlichkeitsstörungen vorherrschen, nicht überraschend. Es ist plausibel, dass beispielsweise Personen mit zwanghaften Persönlichkeitsstörungen aufgrund ihrer Gewissenhaftigkeit die SCL-90-R-Fragebögen häufiger vollständig ausfüllten als ihre Mitpatienten. Außerdem ist es naheliegend, dass Personen mit vermeidend-selbstunsicherer oder dependenter Persönlichkeitsstörung seltener durch einen frühzeitigen Therapieabbruch aus der Stichprobe ausgeschlossen wurden als andere Patienten, da sie beispielsweise das mit einem Therapieabbruch verbundene persönliche Risiko nicht auf sich nehmen wollten oder sie die Entscheidung eines Therapieabbruchs nicht eigenständig treffen konnten.

4.3.2 Diskussion der 1. Forschungsfrage

Wie häufig sich Patienten im Rahmen einer stationären psychotherapeutischen Behandlung verschlechtern und wie häufig Nonresponse auftritt, hängt von dem verwendeten Outcome-Maß ab. Bei der direkten Messung schätzten die Therapeuten die Veränderung der Symptomatik besser ein als die Patienten selbst (Therapeutenperspektive: 2 % Verschlechterung, 10,1 % Nonresponse; Patientenperspektive: 7,5 % Verschlechterung, 12,6 % Nonresponse). In der indirekten Messung waren die Anteile an verschlechterten Patienten bei der Messung mit der SCL-90-R (5,9 %) und dem IIP (5,0 %) ähnlich groß und niedriger bei der Messung mit dem BSS (1,1 %). Von den Nonresponse-Anteilen bei der Messung mit der SCL-90-R (22,9 %) und dem BSS (23,4 %) wich der Anteil bei der Messung mit dem IIP (50 %) deutlich ab. Die Effektstärke war für den BSS-Gesamtwert am größten (d = 1,47) und für den IIP-Gesamtwert am niedrigsten (d = 0,41) (vgl. GSI der SCL-90-R: d = 0,75).

Zu Beginn dieses Kapitels werden die Ergebnisse der direkten Veränderungsmessung diskutiert. Die Häufigkeiten von Patienten mit unveränderter oder verschlechterter Symptomatik dieser Studie lagen alle in einem Bereich, der von anderen Studien schon einmal berichtet wurde: 4,9 % (Junge und Ahrens 1996) bis 24,2 % (Kriebel et al. 2001) für Non-

response und 0,5 % (Rief et al. 1994) bis 13,8 % (Kriebel et al. 2001) für Verschlechterungen (s. 1.2.2 und Tabelle 34).

Tabelle 34: Ergebnisse der direkten Veränderungsmessung in anderen Studien

Quelle	Symptomatik aus Patientenperspektive		Symptomatik aus Therapeutenperspektive	
	unverändert (%)	verschlechtert (%)	unverändert (%)	verschlechtert (%)
Junge und Ahrens 1996	9,2	7,2	4,9	1,8
Kriebel et al. 2001	24,2	13,8	-	-
Nübling et al. 2004	36,4-36,8		-	-
Rief et al. 1994	5,6	4,1	6,3	0,5
Spitzer et al. 2012	18,7		11,1	
Diese Studie	12,6	7,5	10,1	2,0

- : fehlende Angaben

Für Verschlechterung der Symptomatik werden in dieser Studie und der Studie von Junge und Ahrens (1996) vergleichbar niedrige Werte angegeben (s. Tabelle 34). Dort wurde der Therapieerfolg ebenfalls mithilfe einer fünfstufigen Rating-Skala beurteilt, jedoch getrennt nach körperlichen und psychischen Beschwerden (Junge und Ahrens 1996), während in dieser Studie die Veränderung der Symptomatik insgesamt beurteilt wurde. Die dort vorgenommene Trennung in eine Beurteilung der psychischen und der somatischen Beschwerden kann den dort niedrigeren Wert für Nonresponse (4,9 %) erklären, wenn beachtet wird, dass die Therapeuten für deutlich mehr Patienten unveränderte somatische Beschwerden angaben (15,2 % der Patienten (Junge und Ahrens 1996)). In welchem Maß sich die psychisch und somatisch unveränderten Patienten überschneiden, geht aus den Angaben nicht hervor (Junge und Ahrens 1996). Der in dieser Studie ermittelte Nonresponse-Wert aus Sicht der Therapeuten gleicht dem Wert für Misserfolg (unverändert oder verschlechtert) von der Arbeitsgruppe um Spitzer (2012) (s. Tabelle 34). Die große Ähnlichkeit erklärt sich damit, dass sich die Patientenkollektive der beiden Stichproben teilweise überschneiden (Zeitraum 01.01.2005-31.12.2012). Die Anteile an verschlechterten oder unveränderten Patienten sind sowohl aus Therapeuten- als auch aus Patientenperspektive in der Studie von Rief und Kollegen (1994) niedriger. In der genannten Studie wurde eine sechsstufige Skala (wesentlich gebessert, deutlich gebessert, leicht gebessert, unverändert, leicht verschlechtert und deutlich verschlechtert) verwendet (Rief et al. 1994). Diese hat einen Überhang im positiven Bereich. Dadurch, dass „unverändert“ und „leicht verbessert“ um den Mittelwert der Skala herum lagen, besteht die Möglichkeit, dass Patienten, die bei einer fünfstufigen Skala den Mittelwert „unverändert“ gewählt hätten, dort auch „leicht verbes-

sert“ gewählt haben, sodass der Anteil an unveränderten und verschlechterten Personen niedriger ausfiel.

Bei Betrachtung der Patientenperspektive fällt erneut die Ähnlichkeit zwischen den Ergebnissen dieser Arbeit und der Studie von Junge und Ahrens (1996) auf. Der Anteil an Patienten, die selbst eine Verschlechterung angaben, ist in beiden Studien ähnlich (7,5 % in dieser Studie vs. 7,2 % (Junge und Ahrens 1996)). Der Anteil an Patienten, die keine Veränderung feststellten (12,6 %), war nur geringfügig höher als in der Studie von Junge und Ahrens (1996) (9,2 %). Auch die Einschätzung der Patienten in der Studie von Spitzer und Mitarbeitern (2012) stimmten mit der Beurteilung der Patienten in dieser Studie überein. Deutlich höhere Werte für unveränderte und verschlechterte Patienten (ca. 36 %) berichteten Nübling und Mitarbeiter (2004). Die Einschätzung erfolgte dort ebenfalls auf einer fünfstufigen Skala, mit 3 als „unverändert“, jedoch andersherum beschriftet, mit 1 als „deutlich besser“ und 5 als „deutlich schlechter“. Unter Beachtung der unterschiedlichen Polung der Skalen bestätigt ein Vergleich der Mittelwerte (M = 2,38, SD = 0,89 (Nübling et al. 2004) vs. M = 4,04, SD = 0,95), dass die Patienten die Veränderung in der Studie von Nübling und Mitautoren (2004) im Mittel schlechter bewerten. Das Patientenurteil wurde dort jedoch nicht als eine globale Einschätzung der Veränderung der Symptomatik erhoben, sondern die Patienten schätzten subjektiv ihre Veränderung für 13 Items bezüglich körperlicher, psychischer und bewältigungsorientierter Aspekte ein, woraus ein Mittelwert gebildet wurde (Nübling et al. 2004). Patienten mit einem Wert von > 2,65 wurden der Gruppe der nicht gebesserten zugeordnet (Nübling et al. 2004). Das bedeutet, dass auch Patienten, bei denen sich beispielsweise einzelne Aspekte verbessert hatten, während andere unverändert geblieben waren, der Gruppe der nicht gebesserten Patienten zugeordnet wurden. Es bleibt offen, wie ein solcher Patient die Veränderung eingeschätzt hätte, wenn er eine globale Beurteilung hätte abgeben müssen. Je nach Gewichtung der Veränderungen durch die Patienten hätten sich gegebenenfalls andere Anteile ergeben. Durch die reine Mittelwertbildung aus den einzelnen Items, wie sie in der Studie angewandt wurde, werden die Veränderungen aller beurteilten Aspekte als gleich bedeutsam angenommen, ohne dass der Patient eine Gewichtung vornehmen kann. Eine solche subjektive Gewichtung war in der Studie im Fachklinikum Tiefenbrunn durch die Bildung des Globalurteils möglich. Weitere Unterschiede zu der hier vorliegenden Arbeit sind, dass die Arbeitsgruppe um Nübling (2004) nur die Werte der Einjahreskatamnese berichtete und nicht diejenigen zum Zeitpunkt der Entlassung und dass die dort verwendeten Daten im Rahmen einer psychosomatischen Rehabilitation mit deutlich kürzerer Verweildauer erhoben wurden.

Wie Nübling und Kollegen (2004) verwendeten Kriebel und Mitarbeiter (2001) mehrere Items (25 siebenstufige Items) zur direkten Veränderungsmessung aus Patientensicht. Deren Studie (Kriebel et al. 2001) gleicht der Untersuchung von Nübling und Kollegen (2004) auch hinsichtlich der deutlich kürzeren mittleren Behandlungsdauer und der psychosomatischen Rehabilitation als Setting, jedoch mit dem Unterscheid, dass die Erhebung wie im Fachklinikum Tiefenbrunn zum Entlassungszeitpunkt erfolgte. Die Anteile unveränderter

und verschlechterter Patienten lagen somit wie zu erwarten in der Größenordnung der Studie von Nübling und Mitarbeitern (2004). Durch die Unterschiede der Untersuchungsinstrumente und im Studiendesign sind die Ergebnisse der Arbeitsgruppen um Nübling (2004) und um Kriebel (2001) schlechter mit den Werten dieser Studie vergleichbar als die Werte aus Studien, die ein ähnlicheres Studiendesign und ähnlichere Erhebungsarten verwendeten (Junge und Ahrens 1996; Spitzer et al. 2012). Die Ähnlichkeit mit den Ergebnissen der letztgenannten Studien lässt die Werte der direkten Veränderungsmessung dieser Studie insgesamt plausibel erscheinen.

Nachdem sich der vorausgegangene Abschnitt mit den Ergebnissen der direkten Veränderungsmessung beschäftigt hat, werden im Folgenden die Ergebnisse der indirekten Veränderungsmessung vor dem Hintergrund veröffentlichter Studien erläutert. Tabelle 35 gibt einen Überblick über die Ergebnisse der indirekten Veränderungsmessung in anderen Studien. Da in diesen meist die SCL-90-R oder ihre Kurzform, das Brief Symptom Inventory (BSI), als Ergebnismaß verwendet wurde, werden hier zunächst die mit der SCL-90-R erhobenen Befunde zu Nonresponse und zu Verschlechterungen betrachtet, bevor im weiteren Verlauf auch die Ergebnisse des BSS und IIP diskutiert werden.

Der Vergleich der Therapieergebnisse wird dadurch erschwert, dass die Einteilung der Ergebnisgruppen nicht vollständig einheitlich erfolgt. Somit muss bei einem Vergleich beispielsweise beachtet werden, in welcher Form die Gruppe der Pseudogesunden (Spitzer et al. 2018; 2019) (s. 1.3.2) berücksichtigt wurde. Diese wird in einigen Studien unterschieden, während sie in anderen Studien nicht differenziert wird. In Tabelle 35 wird daher gesondert auf das Vorliegen einer zusätzlich erfassten Gruppe von Pseudogesunden hingewiesen.

Tabelle 35: Ergebnisse der indirekten Veränderungsmessung in anderen Studien

Quelle	Test	Erfolg (%)	Non-response (%)	Verschlechterung (%)	Ergänzungen und Anmerkungen
Davies-Osterkamp et al. 1996	SCL	49,6	26,8	9,3	14,4 % Pseudogesunde
Franz et al. 2000	SCL	55	22	6	17 % Pseudogesunde
Geiser et al. 2001	SCL	41,8	44,2	14	
Haase et al. 2008	SCL	45,4[1]	22,4	1,9 [2]	[1] klinisch signifikant verbessert [2] klinisch signifikant verschlechtert
Liebler et al. 2004	SCL	58,6	36,8	4,6	2,3 % Pseudogesunde. Diese sind in den anderen Gruppen enthalten.
Sammet et al. 2004	BSI	60,6	20,6	18,8	
Schauenburg et al. 2001	SCL	59	31	10	
Spitzer et al. 2012	SCL		30,7 [3]		[3] Misserfolg nach Ausschluss der Pseudogesunden. Ca. 22 % Nonresponse und ca. 6 % Verschlechterung vor Ausschluss der Pseudogesunden (10,6 %).
Haase et al. 2008	IIP	44,9	12,2 [4]	6,4 [5]	[4] Nonresponse und statistisch signifikant verschlechtert [5] klinisch signifikant verschlechtert, 36,5 % Pseudogesunde
Oster et al. 2009	multiple [6]		21 [7]		[6] Mehrere Erfolgsmaße zum Befinden, zur Sozialmedizin und zu externen Kosten. Je Bereich jeweils Selbst- und Fremdbeurteilungen, u.a. SCL-90-R, BSS, Fragebogen zur Motivation zur psychosomatisch-psychotherapeutischen Bearbeitung von beruflichen Belastungen (FBTM), direkte Veränderungsmessung. [7] In keinem Bereich gebessert. Als gebessert zählt, wer in der Selbst- und der Fremdeinschätzung eines Bereichs als gebessert gilt.

Tabelle 35 (Fortsetzung): Ergebnisse der indirekten Veränderungsmessung in anderen Studien

Quelle	Test	Erfolg (%)	Non-response (%)	Verschlechterung (%)	Ergänzungen und Anmerkungen
Rudolf et al. 2004	PSKB-Se-R		15-37 [8]		[8] ungebessert, differenziert nach Diagnosen: 15 % bei Angst- und Zwangsstörungen bis 37 % bei somatoformen Störungen. Als „ungebessert“ werden die Patienten bezeichnet, bei denen keine signifikante Besserung oder eine Verschlechterung in einer der vier Symptomskalen des PSKB-Se-R beobachtet wurde.
Spitzer et al. 2012	BSS		26,7 [9]		[9] Misserfolg
Diese Studie	SCL	58,5	23,4	5,9 [10] = 5,0 + 0,9 [11]	12,2 % Pseudogesunde [10] Verschlechterung insgesamt [11] statistisch signifikante + klinisch bedeutsame Verschlechterung
	BSS	69,3	22,9	1,1 [10] = 0,8 + 0,3 [11]	6,8 % Pseudogesunde
	IIP	21,7	50,0	5,0 [10] = 1,9 + 3,0 [11]	23,4 % Pseudogesunde

Die mit Hilfe der SCL-90-R ermittelten Anteile an Patienten mit Nonresponse gleichen den Anteilen, die ebenfalls mit der SCL-90-R bestimmt wurden, in den Studien der Arbeitsgruppen um Haase (2008), Franz (2000), Spitzer (2012), Sammet (2004) und Davies-Osterkamp (1996) (s. Tabelle 35). Es fällt auf, dass in der Studie von Geiser und Mitarbeitern (2001) der Anteil an Nonresponse mit 44,2 % und in der Studie von Liebler und Kollegen (2004) mit 36,8 % deutlich höher ist als in dieser Studie. Dies lässt sich unter anderem damit erklären, dass Geiser und Mitarbeiter (2001) und Liebler und Mitarbeiter (2004) zur Gruppenbildung andere Grenzwerte verwendeten. Während in dieser Studie, beruhend auf den Empfehlungen von Franke und Mitarbeitern (2002), Veränderungen des GSI von ≥ 0,3 als statistisch signifikant bewertet wurden, galten dort erst Veränderungen ≥ 0,43 (Geiser et al. 2001) bzw. ≥ 0,42 (im dysfunktionalen Bereich) (Liebler et al. 2004) als statistisch signifikant, was die hohen Anteile an Nonresponse erklärt. Verstärkend kommt hinzu, dass in beiden Studien (Geiser et al. 2001; Liebler et al. 2004) keine Gruppe von Pseudogesunden differenziert wurde, die in dieser Studie 12,2 % und in den Studien der Arbeitsgruppen um Davies-Osterkamp (1996), Franz (2000) und Spitzer (2012) ebenfalls 10 bis 17 % der Patienten ausmachten (s. Tabelle 35). Die in den letztgenannten Studien als Pseu-

dogesunde klassifizierten Patienten sind dann in den Studien, die diese Gruppe nicht separieren, entsprechend in den übrigen Gruppen (Erfolg, Nonresponse, Verschlechterung) enthalten und erhöhen deren relative Häufigkeit. Beispielsweise berichten Liebler und Kollegen (2004), dass 2,3 % der Patienten bei Aufnahme einen GSI im Bereich einer gesunden Stichprobe aufwiesen, die dann dennoch den übrigen Gruppen zugeordnet wurden. Dass der Anteil der Patienten, die bei Aufnahme einen GSI im Bereich einer gesunden Stichprobe aufwiesen, in der Studie von Liebler und Kollegen (2004) deutlich geringer war als der Anteil an Pseudogesunden in dieser Studie, ist darauf zurückzuführen, dass dort für den GSI ein niedrigerer Cut-off-Wert (c = 0,57) (Liebler et al. 2004), der die funktionalen von den dysfunktionalen Werten trennt, verwendet wurde als hier (c = 0,7).

Nach der Diskussion der Anteile an Nonresponse bei der Messung mit der SCL-90-R folgt nun die Betrachtung der Verschlechterungen. Der mit der SCL-90-R gemessene Anteil an Verschlechterungen (5,9 %, statistisch signifikant oder klinisch bedeutsam) gleicht den berichteten Anteilen in den Studien der Forschungsgruppen um Franz (2000), Liebler (2004) und Spitzer (2012) (s. Tabelle 35). Haase und Mitarbeiter (2008), die denselben Cut-off-Wert (c = 0,7) wie in dieser Studie verwendeten, berichten von 1,9 % klinisch bedeutsam verschlechterten Patienten, was dem Anteil von 0,9 % klinisch bedeutsam verschlechterter Patienten in dieser Studie ähnelt. Diese Anteile an klinisch bedeutsamer Verschlechterung gleichen außerdem jenem Anteil in der Studie von Spitzer und Kollegen (2012). Zusammenfassend sind die in dieser Studie ermittelten Werte für Verschlechterungen also plausibel. Eine deutliche Abweichung von den 5,9 % verschlechterten Patienten lässt sich bei Betrachtung der Verschlechterten (18,8 %) in der Studie von Sammet und Kollegen (2004) feststellen. Sammet und Kollegen (2004), die das Brief Symptom Inventory verwendeten, führten die Abweichung ihres Wertes von den niedrigeren Ergebnissen früherer Psychotherapiestudien darauf zurück, dass der Ausgangswert nicht am Aufnahmetag sondern am Ende der ersten Behandlungswoche erfasst wurde, wodurch Verbesserungen, die in den ersten Tagen einer Therapie häufig seien, nicht gemessen würden (Sammet et al. 2004). Diese Begründung setzt voraus, dass in den ersten Tagen einer Therapie relevante und vorrangig positive Veränderungen stattfinden. Dagegen sprechen jedoch die Ergebnisse von Schauenburg und Kollegen (2001), die einen Abfall der Symptomatik im Durchschnitt erst am Ende der Therapie feststellten und die für die Initialphase einer Therapie nicht nur Patienten mit Verbesserungen (29,6 %), sondern auch mit einer Zunahme der Symptomatik (28,2 %) beobachteten (Schauenburg et al. 2001). Außerdem würde sich die These der Arbeitsgruppe um Sammet (2004) auch in einem niedrigeren Anteil an verbesserten Patienten niederschlagen. Der Anteil an verbesserten Patienten ist jedoch mit 60,6 % hoch (Sammet et al. 2004) im Vergleich zu anderen Studien (s. Tabelle 35). Die Begründung von Sammet und Kollegen (2004) würde durch niedrigere Effektstärken unterstützt werden. Diese werden jedoch nicht erwähnt, sodass kein Vergleich vorgenommen werden kann. Dass die Pseudogesunden nicht differenziert werden, kann den massiven Unterschied zwischen dem Anteil an Verschlechterungen in dieser Studie und in der Studie von Sammet und Kollegen

(2004) auch nicht erklären, da es keinen Anhaltspunkt dafür gibt, dass die Pseudogesunden so stark in der Gruppe der Verschlechterten vertreten sein sollten. Naheliegender ist die Annahme, dass abweichende Grenzwerte zur Gruppenbildung verwendet wurden. Die Kriterien zur Gruppenbildung werden jedoch nicht genannt. Auch in der Studie von Geiser und Kollegen (2001) lag der Anteil an verschlechterten Patienten höher (14 %). Die Studie umfasste jedoch nur eine sehr kleine Stichprobe (n = 43) und berücksichtigte nur Patienten, die auch zwei Jahre nach Therapieende an einer schriftlichen Katamneseerhebung teilgenommen hatten. Geiser und Kollegen (2001) weisen darauf hin, dass sich die Patienten der Katamnesestichprobe und der Gesamtstichprobe, die zusätzlich die Patienten enthält, deren Daten für den Aufnahme-, Entlassungs- und Katamnesezeitpunkt nur unvollständig vorlagen, hinsichtlich der Geschlechterverteilung, des Alters und der Behandlungsdauer nicht unterschieden. Trotzdem kann nicht ausgeschlossen werden, dass sie sich aber hinsichtlich des Therapieerfolgs zum Zeitpunkt des Therapieendes unterschieden. Zusammenfassend stellen die Studien der Arbeitsgruppen um Sammet (2004) und Liebler (2004) somit keinen Widerspruch zu dieser Studie dar. Dies liegt an den unklaren Berechnungsmethoden in der Studie von Sammet und Kollegen (2004) und an der fraglichen Repräsentativität der sehr kleinen Stichprobe und den abweichenden Einschlusskriterien in der Studie von Liebler und Mitarbeitern (2004).

Der BSS wurde in bisherigen Studien zu Misserfolgen in der stationären Psychotherapie zwar bereits eingesetzt (s. Tabelle 36), jedoch verwendeten lediglich Spitzer und Mitarbeiter (2012) den BSS zuvor zur unmittelbaren Einteilung von Erfolgsgruppen. Der Anteil an Misserfolg stimmt aufgrund der Ähnlichkeit der beiden Studien (s.o.) weitestgehend überein (hier: 24,5 % vs. 26,7 % (Spitzer et al. 2012)). Oster und Kollegen (2009) nutzten den BSS auch als Ergebnismaß zur Beurteilung des Befindens. Da sie die Ergebnisse des BSS jedoch mit den Ergebnissen weiterer Tests und mit zusätzlichen Kriterien in Beziehung setzten und Patienten beispielsweise nur als gebessert bewertet wurden, wenn sie in einem Selbst- und einem Fremdbeurteilungsverfahren übereinstimmend als verbessert galten (Oster et al. 2009), können die Ergebnisse schlecht mit denen dieser Studie verglichen werden.

Lediglich die Studie von Haase und Kollegen (2008) verwendet auch das IIP zur Ergebnisgruppenbildung. In der hier vorliegenden Arbeit ist der Anteil an Patienten mit Nonresponse mit 50,0 % deutlich höher als bei Haase und Mitarbeitern (2008), die 12,2 % Nonresponse und statistisch signifikante Verschlechterung berichten. Dies ist vor allem auf die deutlich höhere kritische Differenz zur Gruppenbildung, die im Fachklinikum Tiefenbrunn verwendet wurde, zurückzuführen (0,42 in dieser Studie vs. 0,33 (Haase et al. 2008)). Die verwendeten Cut-off-Werte liegen nah beieinander (1,51 in dieser Studie vs. 1,535 (Haase et al. 2008)). Lediglich 1,7 % bzw. 2 % der Patienten wiesen bei Aufnahme bzw. bei Entlassung in dieser Studie einen Wert zwischen diesen beiden Grenzwerten (also 1,51 bis 1,53) auf, sodass die unterschiedlichen Cut-off-Werte bei der Gruppenbildung kaum ins Gewicht fallen. Unter Berücksichtigung des minimal niedrigeren Cut-off-Werts in dieser Studie und des großen Anteils der Patienten, die auf die Nonresponse-Gruppe entfallen,

erscheinen die Anteile an klinisch signifikanter Verschlechterung mit 3,0 % in dieser Studie vs. 6,4 % in der Studie von Haase und Kollegen (2008) plausibel.

Neben Therapieergebnissen auf Einzelfallebene beschrieben Autoren anderer Psychotherapiestudien häufig die Effektstärken ihrer Therapien (s. Tabelle 36). Diese werden im Folgenden kurz mit den in dieser Studie ermittelten Effektstärken in Beziehung gesetzt.

Tabelle 36: Mittelwerte und Effektstärken der Summenscores der SCL-90-R, des BSS und des IIP in anderen Studien

Studie	SCL-90-R					BSS					IIP				
	Aufnahme		Entlassung		Effektstärke	Aufnahme		Entlassung		Effektstärke	Aufnahme		Entlassung		Effektstärke
	M	SD	M	SD	d	M	SD	M	SD	d	M	SD	M	SD	d
Davies-Osterkamp et al. 1996	1,61	0,72	1,2	0,72	0,57						14,43 1,80 [1]	4,84	12,98 1,62 [1]	3,57	0,3
Franz et al. 2000	1,38	0,66	0,82	0,62	0,84	6,53	1,22	5,44	2,06	0,89	13,64 1,71 [1]	4,35	11,28 1,41 [1]	4,76	0,54
Geiser et al. 2001					0,56										
Haase et al. 2008	ca. 66,6 [2]		55,2 [2]		1,16 [2]										
Junge und Ahrens 1996	1,21	0,62	0,89	0,61	0,52 [3]						1,67	0,47	1,53	0,49	0,30 [3]
Konzag et al. 2004	1,44	1,13	0,95	0,67	0,65	6,6	1,62	5,0	1,81	0,97					
Liebler et al. 2004	1,69	0,62	1,04	0,63											
Nübling et al. 2004	64,5 [2]														
Oster et al. 2009	1,2		0,8		0,6										
Schauenburg et al. 2001	1,38	0,63	0,93	0,75	0,71	7,14	1,47								
Diese Studie	1,43	0,65	0,94	0,67	0,75	6,91	1,76	4,32	1,75	1,47	1,75	0,52	1,54	0,59	0,41

[1] $IIP_{Ges}/8$ ergänzend berechnet

[2] T-Werte

[3] Effektstärke ergänzend berechnet

Die Mittelwerte des GSI und die mit der SCL-90-R gemessene Effektstärke in dieser Studie bewegen sich in dem Rahmen anderer Psychotherapiestudien (s. Tabelle 36). Der Mittel-

wert des GSI, der im Manual der SCL-90-R für stationäre Psychotherapiepatienten angegeben ist (M = 1,20, SD = 0,62) (Franke 2002), differenziert nicht, wie weit die Patienten im Therapieprozess fortgeschritten sind. Dass er zwischen den Mittelwerten bei Aufnahme und bei Entlassung dieser Studie liegt und dass die Werte anderer Studien mit den Mittelwerten und Effektstärken dieser Studie in der Größenordnung übereinstimmen, spricht für die Gültigkeit der gemessenen Werte und lässt annehmen, dass die Ergebnisse – auch die der anderen Forschungsfragen – , die mit der SCL-90-R gemessen wurden, unter Berücksichtigung des Settings verallgemeinerbar sind. Trotz durchgeführter stationärer Psychotherapie erreichten die Patienten zum Entlassungszeitpunkt im Mittel keinen so niedrigen Wert, wie er für die repräsentative Eichstichprobe der SCL-90-R gemessen wurde (M = 0,38, SD = 0,39) (Franke 2002). Dies verdeutlicht, dass die alleinige Messung von Mittelwerten und Effektstärken zur Therapieevaluation nicht zielführend ist, sondern diese, wie in dieser Studie, ergänzend zu der Erhebung von Therapieerfolgsgruppen interpretiert werden können.

Ebenfalls mit Werten früherer Studien stimmen die Mittelwerte und Effektstärken des IIP überein (s. Tabelle 36). Zur besseren Vergleichbarkeit mit dieser Studie und der Studie von Junge und Ahrens (1996) wurden die berichteten Mittelwerte der IIP-Gesamtwerte aus den Studien von Davies-Osterkamp und Kollegen (1996) und Franz und Kollegen (2000) durch 8 dividiert. Die Division durch 8 ergibt sich aus der Anzahl der Items pro Skala. Dass sie in den beiden letztgenannten Studien nicht durchgeführt wurde, ist am ehesten darauf zurückzuführen, dass auf die notwendige Division erst im Rahmen einer Korrektur der zweiten, überarbeiteten und neu normierten Auflage des IIP (Horowitz et al. 2000; Universitätsklinikum Heidelberg - Forschungsstelle für Psychotherapie 2015) hingewiesen wurde.

Die Mittelwerte und Effektstärken, gemessen mit dem BSS, basieren in der Studie der Forschungsgruppe um Franz (2000) nur auf der Erhebung an n = 32 Patienten, da nur eine Klinik im Rahmen der Multi-Center-Studie den BSS einsetzte (Franz et al. 2000). Aufgrund der dadurch eingeschränkten Repräsentativität wird den abweichenden Ergebnissen von Franz und Kollegen (2000) im Vergleich zu dieser Studie keine Bedeutung beigemessen. Da Konzag und Kollegen (2004) nur Patienten mit Persönlichkeitsstörungen berücksichtigen, Schauenburg und Kollegen (2001) nur Werte zum Aufnahmezeitpunkt berichten und Spitzer und Team (2012) den BSS zur Einteilung von Erfolgsgruppen verwenden, aber keine Mittelwerte nennen, ist eine solche Erfassung zu beiden Zeitpunkten mithilfe des BSS bei einer großen Stichprobe stationärer Psychotherapiepatienten bisher im Rahmen der Misserfolgsforschung einmalig und kann nicht mit früher erfassten Messungen verglichen werden.

In den vorausgegangenen Abschnitten sind die Ergebnisse der unterschiedlichen Messverfahren mit den Ergebnissen anderer Studien verglichen worden. Werden diese Ergebnisse nun untereinander verglichen, fällt auf, dass die Therapeutenurteile jeweils positiver ausfielen als die Patientenurteile. So beurteilten die Therapeuten direkt die Veränderung der Symptomatik seltener als unverändert oder verschlechtert als die Patienten (Therapeuten-

perspektive: 2 % Verschlechterung, 10,1 % Nonresponse; Patientenperspektive: 7,5 % Verschlechterung, 12,6 % Nonresponse). Äquivalentes zeigte sich in der Studie von Junge und Ahrens (1996) und den Studien der Arbeitsgruppen um Rief (1994) und Spitzer (2012), wenn die unveränderte und verschlechterte Symptomatik zusammengefasst als Misserfolg betrachtet wurde (s. Tabelle 34). Auch in der indirekten Messung war der Anteil an verschlechterten Patienten unter Verwendung eines Fremdbeurteilungsverfahrens (BSS) niedriger als bei Verwendung von Selbstbeurteilungsverfahren (SCL-90-R und IIP). Die gleiche Beobachtung machten Spitzer und Kollegen (2012) hinsichtlich des BSS im Vergleich zur SCL-90-R (s. Tabelle 35). Das Phänomen, dass die Therapeutenurteile positiver ausfallen, unabhängig von der Art der Veränderungsmessung, lässt sich auch in der Studie von Geiser und Kollegen (2001) finden. Diese beobachteten die Diskrepanz, dass die Patienten, die durch die Selbstbeurteilung in der SCL-90-R als verschlechtert galten, von den Therapeuten in der direkten Veränderungsmessung als unverändert oder verbessert beurteilt wurden. Außerdem wurden die Patienten, die der SCL-90-R zufolge als unverändert galten, zu 70 % von den Therapeuten als leicht verbessert charakterisiert (Geiser et al. 2001). Die Beobachtung einer positiveren Beurteilung seitens der Therapeuten auf der Ebene der Therapie-Erfolgsgruppen wird dadurch ergänzt, dass auch die Effektstärken bei der Fremdbeurteilung höher ausfallen als in der Selbsteinschätzung (BSS d = 1,47, SCL d = 0,75, IIP d = 0,41). Signifikant höhere Effektstärken bei Fremd- als bei Selbstbeurteilungsverfahren wurden auch bei Therapien von Patienten mit Depressionen in einer Metaanalyse gefunden (Cuijpers et al. 2010), was dafür spricht, dass es sich nicht um einen Zufallsbefund handelt. Stieglitz und Hiller (2014) erklären höhere Effektstärken bei der Fremdbeurteilung damit, dass die Therapeuten die Therapieergebnisse überschätzen aufgrund einer positiven Einstellung zu den Therapieergebnissen. Auch dass Therapeuten einen Misserfolg ihrer Patienten gegebenenfalls als Schwäche ihrer eigenen therapeutischen Fähigkeiten empfinden könnten, könnte deren positivere Beurteilung erklären. Es wird somit deutlich, dass Selbst- und Fremdbeurteilungsverfahren einander ergänzend erhoben und ausgewertet werden sollten.

Neben der positiveren Therapeutenbewertung fielen bei dem Vergleich der Therapieergebnisse deutliche Abweichungen der Ergebnisse des IIP von den Ergebnissen des BSS und der SCL-90-R auf. Die niedrige Effektstärke und der hohe Anteil an Nonresponse, die bei der Messung mit dem IIP im Vergleich zu den anderen Erhebungsinstrumenten auffielen, lässt sich teilweise darauf zurückführen, dass das IIP zwar als änderungssensitiv gilt, aber sich die Werte weniger stark im Therapieverlauf verändern als bei anderen Untersuchungsinstrumenten, z. B. der SCL-90-R (Strauß 2003). Hinzu kommt, dass einige Patienten ggf. während der Therapie gelernt haben, ihre zwischenmenschlichen Schwierigkeiten besser wahrzunehmen und zu beschreiben (Franz et al. 2000).

4.3.3 Diskussion der 2. Forschungsfrage

In diesem Kapitel werden zunächst die Ergebnisse der Untersuchung, ob sich Männer und Frauen in der Auftretenshäufigkeit von Misserfolg unterscheiden, vor dem Hintergrund bereits bestehender Studien bewertet. Darauf folgt die Diskussion des Befundes, dass die Frauen in der indirekten Veränderungsmessung im Mittel höhere Symptombelastungen, interpersonale Probleme und Beeinträchtigungsschweren aufwiesen.

Männer und Frauen unterschieden sich nicht hinsichtlich des Therapieerfolges. Eine Ausnahme bildete die Messung mithilfe der SCL-90-R. Bei dieser wiesen Männer häufiger Misserfolge auf als Frauen. Das Ausmaß, wie stark sich Männer und Frauen hinsichtlich des Auftretens von Nonresponse und Verschlechterungen in der Messung mit der SCL-90-R unterschieden, war jedoch minimal. Mit standardisierten Residuen < 2 und Cramers V < 0,1 kann von keinem relevanten Zusammenhang gesprochen werden. Dieses unterstützen die Ergebnisse der Varianzanalyse, die ergab, dass bei der Messung mit der SCL-90-R nur 2 % der Variation des GSI durch die Gruppenzugehörigkeit zu den Männern oder Frauen und nur 0,2 % der Variation durch die Wechselwirkung zwischen der Geschlechtszugehörigkeit und dem Messzeitpunkt erklärbar sind. Die Anteile sind bei den Messungen mit dem BSS und dem IIP noch niedriger. Vor diesem Hintergrund lässt sich zusammenfassen, dass in dieser Studie unter Verwendung multimodaler Diagnostik an einer großen Stichprobe kein für den Alltag relevanter Unterschied zwischen Männern und Frauen hinsichtlich des Auftretens von Nonresponse oder Verschlechterung festgestellt werden konnte.

Die Ergebnisse dieser Studie bestätigen somit die Ergebnisse der Studien von Bassler (1995) und den Arbeitsgruppen um Davies-Osterkamp (1996), Geyer (2003), Kriebel (2001) und Konzag (2004), die ebenfalls keinen signifikanten oder prognostisch relevanten Einfluss des Geschlechts auf das Therapieergebnis fanden. Dass in der Studie von Geiser und Kollegen (2001) keine Männer Misserfolge aufwiesen, ist der niedrigen Stichprobengröße geschuldet (n = 43, davon 12 Männer) und widerspricht somit nicht den Befunden dieser und der oben genannten Studien. Franz und Mitarbeiter (2000) beschrieben in ihren Ergebnissen einen minimal höheren Anteil an Nonresponse und Verschlechterungen bei Männern als bei Frauen (29 % der Männer vs. 27 % der Frauen) bei der Messung mit der SCL-90-R. Trotzdem fassten sie in ihrer Diskussion zusammen, dass „keine statistisch signifikanten Geschlechtseffekte hinsichtlich symptomatischer Besserung und Veränderung interpersoneller Problembelastung“ (Franz et al. 2000: 255) bestünden.

Der Informationsgewinn durch die zweite Forschungsfrage dieser Studie im Gegensatz zu den zuvor genannten Studien besteht darin, dass erstmalig eine so große Stichprobe untersucht wurde, was zu einer besseren Repräsentativität und Verallgemeinerbarkeit der Ergebnisse führt (s. Tabelle 31). Außerdem ist dies die erste Studie, die das Vorliegen von Geschlechtsunterschieden unter Verwendung multimodaler Diagnostik untersucht. Bassler und Kollegen (1995) verwendeten allein eine Kombination aus der Erfolgszufriedenheit der

Patienten und der Einschätzung der Verbesserung aus Therapeutensicht als Outcome-Maß. Die Arbeitsgruppen um Davies-Osterkamp (1996), Geiser (2001) und Konzag (2004) beschränkten sich auf die Ergebnisse, die die SCL-90-R ergab, und Kriebel und Mitarbeiter (2001) nutzten lediglich die Kurzform des Veränderungsfragebogens des Erlebens und Verhaltens (VEV-K) zur Evaluation. In der Studie von Geyer und Mitarbeitern (2003) ist unklar, wie die Zuordnung der Patienten zur verbesserten oder unveränderten Patientengruppe erfolgte. Franz und Kollegen (2000) untersuchten zwar die Effektstärken der SCL-90-R und des IIP auf Unterschiede zwischen Männern und Frauen, aber auf der Ebene der Therapieerfolgsgruppen testeten sie nur die Werte der SCL-90-R auf Geschlechtsunterschiede. Kritisch anzumerken ist, dass nicht alle der oben genannten Studien analysierten, ob ein Unterschied zwischen Männern und Frauen hinsichtlich der Zugehörigkeit zur Gruppe der Patienten mit Nonresponse oder Verschlechterungen vorlag. So fußt die Aussage von Konzag und Kollegen (2004), dass keine relevanten Zusammenhänge für das Geschlecht und den Therapieerfolg bestünden, darauf, dass sie keine signifikante Korrelation zwischen dem Geschlecht und der Differenz des GSI zum Zeitpunkt der Aufnahme und der Entlassung fanden (Konzag et al. 2004). Diese Studie erlaubt somit erstmalig die Aussage, dass sich Männer und Frauen in der direkten Veränderungsmessung und der Messung mit dem BSS und dem IIP in der Auftretenshäufigkeit von Erfolg und Misserfolg nicht unterscheiden.

Wie oben angekündigt folgt an dieser Stelle die Diskussion des Befundes, dass die Frauen in der indirekten Veränderungsmessung im Mittel höhere Symptombelastungen, interpersonale Probleme und Beeinträchtigungsschweren aufwiesen als die Männer. Dass die Skalensummenwerte bei den Frauen höher als bei den Männern waren, ohne dass sie sich in den Therapieergebnisgruppen unterschieden, kann teilweise darauf zurückzuführen sein, dass Frauen ggf. differenziertere und komplexere Krankheitskonzepte mit höherer Symptomaufmerksamkeit und bewussterem Körpererleben besitzen (Möller-Leimkühler 2005; Riecher-Rössler 2000), sodass sie höhere Werte angeben. Außerdem wird diskutiert, dass Frauen sich ihrer Emotionen bewusster sind und diese offener berichten, da sie dies weniger als stigmatisierend betrachten als Männer (Mirowsky und Ross 1995). Es besteht also aufgrund von Rollenverständnissen eine größere Krankheitsbereitschaft und Klagsamkeit bei Frauen als bei Männern (Möller-Leimkühler 2005). Diesen gegenüber stehen Dissimulationstendenzen von Männern (Mirowsky und Ross 1995). Ein weiterer Grund kann aber auch eine tatsächlich stärkere Belastung der Frauen sein. So fanden Mirowsky und Ross (1995) in einer Studie an 1282 Frauen und 749 Männern, dass Frauen 30 % häufiger Sorgen hatten als Männer. Eine weitere Ursache für höhere Belastungswerte von Frauen als von Männern, die in der Literatur diskutiert wird, bezieht sich auf die verwendeten Tests. Unter der Berücksichtigung, dass Frauen mit anderen Emotionen auf Stress reagieren als Männer, weisen Frauen der Theorie nach höhere Werte als Männer auf, wenn das Erhebungsinstrument mehr Symptome bzw. Emotionen abfragt, die häufiger bei Frauen als bei Männern auftreten, z. B. Traurigkeit / Depression (Mirowsky und Ross 1995). Diese Ursache

ist für diese Arbeit jedoch nicht naheliegend, da mit der SCL-90-R und dem IIP Fragebögen verwendet wurden, die sich durch ihre Vielzahl an Skalen nicht auf Aspekte beschränken, die häufiger von Frauen erlebt werden, sondern die auch Bereiche abfragen, die als typisch für Männer gelten. Mirowsky und Ross (1995) nennen dafür expressives Verhalten, Wut oder Feindseligkeit als Beispiele.

Entsprechend der höheren GSI-Werte der Frauen existieren auch nach Geschlecht getrennte Cut-off-Werte für den GSI. Für stationäre Psychotherapiepatienten wird ein Cut-off für Frauen von c = 0,77 und für Männer von c = 0,57 angegeben (Franke 2002). In dieser Arbeit wurde dennoch der alters- und geschlechtsunabhängige Cut-off-Wert c = 0,7 verwendet. Inwiefern sich die Zuordnung zu den Ergebnisgruppen durch die getrennten Cut-off-Werte ändern würde, bleibt offen. Für die Verwendung dieses allgemeinen Grenzwerts spricht die einfache und transparente Einsetzbarkeit. Diese ginge verloren, wenn die geschlechtsabhängigen Cut-off-Werte und die nach Hauptdiagnosen differenzierten Cut-off-Werte (Franke 2002) verwendet würden. Hinzu käme, dass entschieden werden müsste, welche Differenzierung die relevanteste ist, da beispielsweise die Cut-off-Werte, die nach Diagnosen differenzieren, noch nicht das Geschlecht berücksichtigen.

4.3.4 Diskussion der 3. Forschungsfrage

In diesem Kapitel werden zuerst die Ergebnisse der Prüfung auf Unterschiede zwischen Patienten mit und ohne Persönlichkeitsstörungen und im Anschluss die Resultate des Vergleichs von Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen erörtert. Im Anschluss folgt eine kritische Betrachtung der Diagnostik der Persönlichkeitsstörungen.

In der direkten Messung wurde die Veränderung der Symptomatik sowohl von den Therapeuten als auch den Patienten selbst für jene mit Persönlichkeitsstörungen schlechter eingeschätzt als für Patienten ohne Persönlichkeitsstörungen. In der indirekten Veränderungsmessung unterschieden sich diese beiden Patientengruppen jedoch nicht hinsichtlich der Häufigkeit von Nonresponse und Verschlechterung. In den verwendeten Untersuchungsinstrumenten zur indirekten Messung wiesen Patienten mit Persönlichkeitsstörungen im Mittel höhere Werte auf als Patienten ohne Persönlichkeitsstörungen.

Es stellt sich die Frage, warum sich die Personengruppen in der direkten Veränderungsmessung signifikant unterschieden, während in der indirekten Messung kein signifikanter Unterschied im Auftreten von Misserfolg oder Erfolg festgestellt wurde. Ein zu beachtender Aspekt ist, dass im t-Test bei der direkten Messung die Mittelwerte der Patientengruppen verglichen wurden. Zum signifikanten Ergebnis trug also auch mit bei, dass in der Selbsteinschätzung der Anteil an Patienten mit Persönlichkeitsstörungen, die ihre Symptomatik als deutlich gebessert beurteilten (bestmögliche Antwortmöglichkeit), um 6,5 % niedriger lag als der Anteil an Patienten ohne Persönlichkeitsstörungen mit derselben Beurteilung. Die Therapeuten schätzten die Patienten mit Persönlichkeitsstörungen zu 5,1 % seltener als deutlich gebessert ein als die Patienten ohne Persönlichkeitsstörungen. Diese Un-

terschiede im positiven Veränderungserleben waren deutlich größer als bei den negativen Erfahrungen. Die Therapeutenurteile, wie viele Patienten sich verschlechterten bzw. unverändert blieben, wichen für Patienten mit Persönlichkeitsstörungen um 0,7 % bzw. 0,6 % von den Anteilen für Patienten ohne Persönlichkeitsstörungen ab. Die Patienten mit Persönlichkeitsstörungen wiesen in der Selbsteinschätzung 1,3 % mehr Verschlechterungen und 1,3 % mehr unveränderte Symptomatik als die Patienten ohne Persönlichkeitsstörungen auf. Bei der indirekten Messung blieben die Unterschiede innerhalb des positiven Bereichs hingegen unbeachtet, was erklären kann, warum in der direkten Messung der Unterschied zwischen den Mittelwerten signifikant war, während die Gruppenunterschiede in der indirekten Messung nicht signifikant waren.

Die vorliegenden Ergebnisse sind nicht mit Ergebnissen anderer Studien zu vergleichen, da zu dieser Fragestellung keine Publikationen vorliegen. Franz und Kollegen (2000) gaben zwar 50 % Nonresponse für Patienten mit Borderline-Persönlichkeitsstörungen an, aber ihre Aussage bezog sich lediglich auf vier Patienten, was keiner repräsentativen Stichprobe entspricht. Rudolf und Mitarbeiter (2004) verglichen zwar nicht die Patientengruppe mit Persönlichkeitsstörungen mit derjenigen ohne Persönlichkeitsstörungen, aber sie stellten den Anteil an Patienten, bei denen keine signifikante Besserung oder eine Verschlechterung in einer der vier PSKB-Se-R-Symptomskalen vorlag, heraus. Dieser Anteil lag für Patienten mit Persönlichkeitsstörungen bei 19-22 %. Ähnliche Werte fanden sie für Essstörungen (21 %) und affektive Störungen (23 %) (Rudolf et al. 2004). Dies ließ vermuten, dass die Anteile an Nonresponse und Verschlechterung bei Patienten mit Persönlichkeitsstörungen nicht deutlich höher als bei Patienten mit anderen psychiatrischen Diagnosen sind.

Die Patienten mit Cluster-B- und -C-Persönlichkeitsstörungen unterschieden sich nicht hinsichtlich der Häufigkeit von Therapiemisserfolg. Lediglich bei der Messung mit der SCL-90-R wiesen die Patienten mit Custer-B-Persönlichkeitsstörungen höhere Symptombelastungen auf als diejenigen mit Cluster C-Persönlichkeitsstörungen. Es bestand jedoch keine Wechselwirkung zwischen dem Persönlichkeitsstörungscluster und dem Messzeitpunkt, was bedeutet, dass das Cluster das Therapieergebnis nicht beeinflusst.

Bisher wurden in anderen Studien häufig die Ergebnisse einzelner Störungsbilder untersucht und selten mehrere oder die Persönlichkeitsstörungscluster verglichen. In Anlehnung an die Ergebnisse von Mohr (1995), der höhere Misserfolgsraten bei Patienten mit Borderline-Störungen benannte, hätte ein Trend dahingehend erwartet werden können, dass bei Patienten mit Cluster B-Persönlichkeitsstörungen, zu denen die Patienten mit Borderline-Persönlichkeitsstörung gehören, mehr Misserfolge auftreten. Spitzer und Mitarbeiter (2012) verglichen die Cluster miteinander, wobei in ihrer Untersuchung die Patienten mit Cluster-A-Persönlichkeitsstörung häufiger erfolgreich und die Patienten mit Cluster-B- und kombinierten Persönlichkeitsstörungen häufiger erfolglos behandelt wurden (Spitzer et al. 2012). Die Unterschiede waren dabei meist zwischen Cluster-A- und Cluster-B- oder den kombinierten Persönlichkeitsstörungen am größten, aber trotzdem noch gering (Spitzer et al.

2012). Der Trend, dass die Patienten mit Cluster-B-Persönlichkeitsstörung mehr Misserfolge bei der Beurteilung mit Hilfe der SCL-90-R und des BSS aufwiesen (Spitzer et al. 2012), war in dieser Studie zwar reproduzierbar, jedoch nicht signifikant. Dies kann auch daran liegen, dass in dieser Studie der Unterschied zum Cluster A nicht untersucht wurde. Für die Messung mit dem IIP galt dieser Trend bereits nicht mehr, sodass insgesamt aus den Ergebnissen dieser Arbeit keine Abhängigkeit des Therapieergebnisses vom Persönlichkeitsstörungscluster- B oder -C abgeleitet werden kann. Da nur 1,7 % der Patienten der Stichprobe eine Cluster-A-Persönlichkeitsstörung aufwiesen, konnten Unterschiede zu diesen Patienten hier nicht untersucht werden. Dies könnte Gegenstand weiterer Studien sein.

In der aktuell abgelaufenen AWMF-Leitlinie zu Persönlichkeitsstörungen wird zur Diagnostik einer Persönlichkeitsstörung die Durchführung eines halbstrukturierten klinischen Interviews empfohlen (S2-Leitlinie Persönlichkeitsstörungen 2009). Die Diagnose der Persönlichkeitsstörung wurde in dieser Arbeit jedoch ohne ein solches Interview, allein durch die klinische Beurteilung gestellt. Dies reduziert die Aussagekraft der Ergebnisse, die die Patientengruppen mit Persönlichkeitsstörungen betreffen. Ein strukturiertes Verfahren zur Diagnostik, z. B. das Strukturierte Klinische Interview zur Diagnostik von Persönlichkeitsstörungen (SKID) oder die International Personality Disorder Examination (IPDE) (S2-Leitlinie Persönlichkeitsstörungen 2009; Fydrich 2008; Habermeyer und Herpertz 2009), wäre somit für weitere Studien wünschenswert. Bezüglich der Reliabilität der Diagnostik von Persönlichkeitsstörungen fasste Fydrich (2008) zusammen, dass die Einschätzung, ob eine Persönlichkeitsstörung vorhanden ist, mit höherer Reliabilität gelingt als eine differenzierte Unterscheidung der einzelnen Persönlichkeitsstörungen. Somit ist eine geringere Einschränkung der Aussagekraft bei den Analysen, die die Patienten mit Persönlichkeitsstörungen mit jenen ohne Persönlichkeitsstörungen vergleichen, anzunehmen als bei den Untersuchungen der einzelnen Persönlichkeitsstörungscluster. Fydrich (2008) führte die höhere Reliabilität strukturierter klinischer Interviews u.a. darauf zurück, dass die Interviewer die diagnostischen Regeln im Rahmen von Studien gewissenhafter berücksichtigen. Da im Fachklinikum Tiefenbrunn wiederholt Studien durchgeführt wurden (z. B. Spitzer et al. 2012; 2015; 2018; 2019), könnte der Argumentation von Fydrich (2008) folgend eine höhere Reliabilität der Diagnostik vorliegen als in Kliniken, in denen keine Studien durchgeführt werden. Somit erscheint, auch unter der Berücksichtigung der Effizienz, das Fehlen eines strukturierten Interviews vertretbar.

4.3.5 Soziodemographische Parameter als Confounder

Im Rahmen der Analyse, ob soziodemographische Unterschiede zwischen den untersuchten Gruppen die Ergebnisse der zweiten und dritten Forschungsfrage konfundierend beeinflussten, konnte gezeigt werden, dass sich Männer und Frauen und Patienten mit und ohne Persönlichkeitsstörungen hinsichtlich einiger soziodemographischer Aspekte statistisch signifikant unterschieden (s. 3.5). Betrachtet man jedoch beispielsweise die bestehenden soziodemographischen Unterschiede zwischen den Männern und Frauen genauer, fällt

auf, dass diese keine sinnvoll interpretierbaren inhaltlichen Trends abbilden. Beispielsweise wiesen die männlichen Patienten einerseits häufiger Hauptschulabschlüsse auf, andererseits schlossen sie die Schule häufiger mit einem Abitur ab als die behandelten Frauen, die häufiger Realschulabschlüsse besaßen, sodass keine Gruppe als schulisch höher gebildet gelten konnte. Auch die Höhe der Berufsabschlüsse ergab keinen relevanten Trend, denn die Männer hatten einerseits häufiger keinen Abschluss, andererseits auch häufiger einen universitären oder Fachhochschulabschluss als die Frauen. Ein Niveauunterschied der Berufe der Männer und Frauen ließ sich ebenfalls nicht feststellen. So waren die behandelten Männer häufiger Arbeiter und Facharbeiter und höhere Beamte sowie Angestellte als die Frauen, dafür waren die Frauen häufiger mittlere Beamte und Angestellte. Eine Beeinflussung des Therapieerfolgs war basierend auf dieser inhaltlichen Interpretation bei fehlenden Trends nicht zu erwarten. Zur Prüfung, wie groß der Einfluss der soziodemographischen Aspekte tatsächlich auf die Auftretenswahrscheinlichkeit von Misserfolg oder Erfolg war, wurde eine logistische Regression durchgeführt. In das Modell dieser logistischen Regression wurden die Berufstätigkeit, der höchste Berufsabschluss und das Alter nicht aufgenommen. Dies verdeutlicht, dass deren Ausprägungen keine Bedeutung für das Auftreten von Erfolg oder Misserfolg haben. Der Familienstand, der höchste Schulabschluss, die Beschäftigung vor Aufnahme, das Vorliegen einer Berentung und das Geschlecht wurden zwar als Prädiktoren in das Modell aufgenommen und hatten einen signifikanten Einfluss auf das Auftreten von Erfolg oder Misserfolg, die Modellgüte war jedoch sehr niedrig. Lediglich 2,3 % der Varianz der abhängigen Variable, also dem Auftreten von Misserfolg, konnten durch alle hier untersuchten soziodemographischen Faktoren zusammen erklärt werden. Dass die soziodemographischen Faktoren somit maximal einen kleinen Einfluss auf das Auftreten von Erfolg oder Misserfolg haben, deckt sich mit den bereits veröffentlichten Studien, die häufig kontroverse Ergebnisse zum Einfluss soziodemographischer Parameter auf das Therapieergebnis darlegten (s. 1.2.3). Durch den höchstens minimalen Einfluss der soziodemographischen Parameter auf das Therapieergebnis können somit mögliche Unterschiede oder Gemeinsamkeiten im Therapieergebnis vorrangig auf die untersuchte Variable selbst und nicht maßgeblich auf die soziodemographischen Aspekte zurückgeführt werden. Streng genommen gilt diese Aussage nur für die Therapieergebnisse, die mit Hilfe der SCL-90-R ermittelt wurden. Da in Anlehnung an die bestehenden Arbeiten zum Einfluss soziodemographischer Parameter auf Misserfolge in der stationären Psychotherapie (s. 1.2.3) und nach obiger logistischer Regression kein Anhalt für einen starken Einfluss soziodemographischer Faktoren auf das Auftreten von Misserfolgen vorliegt, wurde auf weitere logistische Regressionen, die die Therapieergebnisse basierend auf dem IIP und dem BSS verwenden, verzichtet.

4.4 Limitationen der Untersuchung

Einige kritisch zu betrachtende Aspekte dieser Arbeit wurden in den vorausgegangenen Kapiteln bereits diskutiert. Im Folgenden werden weitere Limitationen zusammengefasst dargestellt.

In Hinblick auf das Studiendesign ist das retrospektive Design kritisch zu bewerten. Dadurch konnte beispielsweise auf die Rate fehlender Daten kein Einfluss genommen werden. Des Weiteren waren die Maßnahmen zur Sicherung der internen Validität begrenzt. So wurden vor der Erhebung keine Beobachtungen zu mehreren Zeitpunkten oder kein Vergleich mit anderen Gruppen durchgeführt. Außerdem wurde bei der Untersuchung der Einflussnahme von Interaktionen oder Moderatorvariablen nur der Einfluss soziodemographischer Aspekte auf das Therapieergebnis untersucht und es wurden keine Ergebnisse unter nicht-randomisierten Vergleichsbedingungen analysiert. Trotz der großen Stichprobe ist die Repräsentativität und Verallgemeinerbarkeit der Ergebnisse durch die Anwendung von Ein- und Ausschlusskriterien und dadurch, dass die Erhebung nur in einer Klinik stattfand und nicht als Multi-Center-Studie erfolgte, geringgradig eingeschränkt.

Bezüglich der Bedeutsamkeit der Ergebnisse sind die im Folgenden genannten Aspekte limitierend. Die Aussagekraft der Befunde, die sich auf Patienten mit Persönlichkeitsstörungen beziehen, wird dadurch reduziert, dass diese Diagnose klinisch gestellt wurde und nicht mittels eines standardisierten Interviews. Berücksichtigt werden muss, dass aufgrund der großen Stichprobengröße auch kleine Effekte als statistisch signifikant gemessen werden. Dadurch ist es möglich, dass sich ein signifikanter Unterschied darstellen lässt, der jedoch keine praktische Bedeutung hat. Somit sollten die Ergebnisse auf ihre klinische Relevanz überprüft werden. Außerdem ist kritisch anzumerken, dass bei der indirekten Veränderungsmessung nur die Skalensummenwerte ausgewertet wurden und nicht die einzelnen Skalen. Inwiefern dort Unterschiede bestehen und inwiefern sich diese auf das Therapieergebnis auswirken oder wie groß die Anteile an Nonresponse oder Verschlechterung sind, wenn die Ergebnisse der einzelnen Skalen zur Gruppeneinteilung verwendet werden, bleibt unbeantwortet. Die Aussagekraft der Ergebnisse ist zusätzlich dadurch begrenzt, dass allein die Ergebnisse zum Entlassungszeitpunkt und nicht zusätzlich zu einem Katamnesezeitpunkt untersucht wurden. Somit erlauben die Daten keine Vorhersage, ob das erreichte Therapieergebnis bis zu einem späteren Zeitpunkt stabil und dadurch für die Patienten über den Zeitraum des stationären Aufenthalts hinaus relevant bleibt. Besonders dem IIP wird zugesprochen, dass größere Effekte erst nach längeren Katamnesezeiträumen deutlich würden (Strauß 2003), sodass die Aussagekraft der Ergebnisse des IIP zum Entlassungszeitpunkt limitiert ist.

4.5 Schlussfolgerungen und Ausblick

Von den zuvor genannten Limitationen abgesehen kann diese Arbeit im Vergleich zu anderen Studien zu Misserfolgen in der stationären Psychotherapie durch eine hohe Repräsentativität auf Grund einer großen Stichprobe bei einem langen Studienzeitraum (10 Jahre) und durch die Verwendung objektiver, reliabler und gut validierter Tests überzeugen. Ihre Ergebnisse sind somit unter Berücksichtigung des Settings verallgemeinerbar. Die Arbeit zeichnet sich durch die umfassende Kombination von Selbst- und Fremdbeurteilungsverfahren sowohl in der direkten als auch indirekten Veränderungsmessung im Rahmen multimodaler Diagnostik aus. Außerdem erfolgte die Auswertung auf Einzelfallebene, woraus sich ein Informationsgewinn gegenüber früheren Studien ergibt, die sich häufig auf den Bericht von Effektstärken beschränkten. Therapiemisserfolge wurden in diesem Rahmen erstmalig auch hinsichtlich interpersoneller Aspekte definiert und untersucht. Zum aktuellen Zeitpunkt liegt keine vom Umfang vergleichbare Studie im Bereich der Misserfolgsforschung vor.

Die Studie erlaubt in Einklang mit den bestehenden Studien die Feststellung, dass Nonresponse und Verschlechterungen in der stationären Psychotherapie häufig sind. Somit ist zum Wohle der Patienten, ihres Umfelds und des Gesundheitssystems eine breit aufgestellte Fortsetzung der Ursachenforschung mit dem Ziel der Prävention von Misserfolgen, beispielsweise durch eine frühzeitige Therapieanpassung, notwendig. Wünschenswert wäre, wenn diese Studie dabei hilft, andere Kliniken ebenfalls zur Misserfolgsforschung zu ermutigen. Sie könnte einen Beitrag dazu leisten, indem sie auf die hohe Zahl an Misserfolgen aufmerksam macht, somit für dieses Thema sensibilisiert und die Enttabuisierung von Therapiemisserfolgen fördert. Ein offener Umgang der Kliniken mit Misserfolgsfällen setzt voraus, dass Therapiemisserfolg nicht als persönliches Versagen eines einzelnen Behandlers oder eines therapeutischen Teams verstanden wird.

Im Rahmen dieser Arbeit konnten – unter Berücksichtigung sämtlicher Testergebnisse – weder das Geschlecht noch das Vorliegen einer Persönlichkeitsstörung oder die einzelnen Persönlichkeitsstörungscluster als eindeutige Risikofaktoren für Misserfolge ausgemacht werden. Lediglich zwei Messverfahren erbrachten hiervon abweichende Ergebnisse. Der höhere Anteil an Misserfolgen bei Männern bei der Messung mit der SCL-90-R war jedoch lediglich minimal und der signifikante Unterschied zwischen Personen mit und ohne Persönlichkeitsstörungen in der direkten Veränderungsmessung ließ sich am ehesten auf die Messmethode (Unterschiede vor allem im Erfolgsbereich der Ratingskala) zurückführen. Dadurch relativiert sich die Bedeutsamkeit dieser abweichenden Ergebnisse für das Auftreten von Misserfolgen und die Annahme, dass keine Unterschiede hinsichtlich der Häufigkeit von Therapiemisserfolgen zwischen den untersuchten Gruppen bestehen, scheint den Ergebnissen dieser Arbeit zufolge insgesamt gerechtfertigt. Hieraus lässt sich schlussfolgern, dass der Schwerpunkt weiterer Forschung zunächst auf der Suche nach anderen möglichen Ursachen für Therapiemisserfolg liegen sollte. Die weitere Analyse von Unterschie-

den des Geschlechts und des Vorliegens von Persönlichkeitsstörungen scheint nach den o. g. Befunden wenig zielführend. Ergänzend sollte beachtet werden, dass in dieser Arbeit ausschließlich der Einfluss von Kriterien, die zum Aufnahmezeitpunkt bereits bestanden, auf das Auftreten von Misserfolg untersucht wurde. Darüber hinaus kann eine Untersuchung von Prozessmerkmalen interessant sein. Einigen Autoren zufolge haben Prozessmerkmale einen größeren Einfluss auf das Therapieergebnis als die Eingangskriterien (Kriebel et al. 2001; Paar 1997).

Des Weiteren fiel in dieser Arbeit eine Diskordanz zwischen den ermittelten Häufigkeiten von Misserfolgen in Abhängigkeit von der Erhebungsart auf. Trotz der Interpretationsschwierigkeiten, die sich durch mehrere Erhebungsverfahren ergeben können, sollte auf eine multimodale Diagnostik in zukünftigen Studien nicht verzichtet werden. Eine solche ist unabdingbar, da bisher keine Standards zur Messung von Therapiemisserfolgen auf Einzelfallebene bestehen und sich die Ergebnisse der einzelnen Verfahren in ihrem Informationsgehalt ergänzen. Eine Entwicklung von standardisierten Verfahren zur Messung von Therapiemisserfolg wäre für die weitere Misserfolgsforschung hilfreich.

Insgesamt erlaubt diese Studie Aussagen zu Nonresponse und Verschlechterungen zum Zeitpunkt der Entlassung. Zur Stabilität der Ergebnisse im zeitlichen Verlauf nach der Entlassung der Patienten können jedoch leider keine Aussagen getroffen werden, da keine Nacherhebung durchgeführt wurde. Hier besteht weiterer Forschungsbedarf. Beispielsweise könnte unter Verwendung derselben Tests eine prospektive naturalistische Studie um eine Erhebung im zeitlichen Verlauf nach der Entlassung ergänzt werden.

Darüber hinaus bleibt aktuell unklar, inwiefern die Therapiemisserfolge kausal mit der Behandlung zusammenhängen oder durch den Spontanverlauf der Erkrankung oder andere Einflüsse bedingt sind. Wahrscheinlich wird dies jedoch zunächst auch bei weiteren naturalistischen oder experimentellen Studien, u.a. in Ermangelung einer ethisch vertretbaren und im Rahmen von Studien beobachtbaren Kontrollgruppe, unbeantwortet bleiben.

5 Zusammenfassung

Die Forschungsbestrebungen zu Misserfolgen in der stationären Psychotherapie in Deutschland waren bisher überschaubar und von methodischen Problemen geprägt. Entsprechend fehlen multimodal erhobene Angaben bezüglich der Häufigkeit von Nonresponse und Verschlechterungen sowie hinsichtlich möglicher Faktoren, die das Auftreten von Nonresponse und Verschlechterungen bedingen.

Um diese Leerstellen zu füllen, wurde in dieser Arbeit untersucht, wie häufig Nonresponse und Verschlechterungen vorkommen und ob sich Männer und Frauen oder Personen mit und ohne Persönlichkeitsstörungen oder Personen mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen in der Auftretenshäufigkeit von Misserfolg (Nonresponse und Verschlechterung) unterscheiden.

Hierzu erfolgte eine retrospektive Datenanalyse der Routinediagnostik des Fachklinikums Tiefenbrunn. Es erfüllten 3862 Patienten (59,6 %) der 6481 Patienten, die im Zeitraum vom 01.01.2005 bis zum 31.12.2014 im Asklepios Fachklinikum Tiefenbrunn eine psychodynamisch orientierte stationäre Behandlung erhielten, die Einschlusskriterien. Zu diesen Einschlusskriterien zählten eine Mindestbehandlungsdauer von sechs Wochen, ein Mindestalter von 18 Jahren und die Dokumentation des Global Severity Index der SCL-90-R des Aufnahme- und Entlassungszeitpunkts. Unter Verwendung der Symptom-Checkliste SCL-90-R und des Inventars zur Erfassung interpersonaler Probleme IIP beurteilten die Patienten ihre Symptombelastung und interpersonalen Probleme (indirekte Veränderungsmessung). Die Therapeuten schätzten die Schwere der Erkrankung ihrer Patienten mithilfe des Beeinträchtigungs-Schwere-Scores BSS ein (indirekte Veränderungsmessung). Ergänzend wurde die Veränderung der Symptomatik durch den Patienten und den Therapeuten nach der Therapie eingeschätzt (direkte Veränderungsmessung).

Die Analyse der Daten zeigte, dass sich 1,9-7,5 % der Patienten verschlechterten und 10,1-50,0 % der Patienten als Nonresponder einzustufen waren, abhängig von der Art der Verlaufsmessung, der Perspektive und dem angewendeten Untersuchungsverfahren.

Es konnte kein eindeutiger Unterschied in der Auftretenshäufigkeit von Nonresponse und Verschlechterung zwischen Männern und Frauen festgestellt werden. Lediglich bei der Messung mithilfe der SCL-90-R wiesen Männer häufiger Misserfolge als Frauen auf.

Darüber hinaus zeigte sich beim Vergleich der Patienten mit und ohne Persönlichkeitsstörungen im Rahmen der indirekten Veränderungsmessung kein Unterschied im Hinblick auf das Auftreten von Therapiemisserfolg. Die direkte Verlaufsmessung ergab, dass sich Patienten mit Persönlichkeitsstörungen häufiger als verschlechtert oder unverändert einschätzten als jene ohne Persönlichkeitsstörungen. Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörungen unterschieden sich weder in der direkten noch in der indirekten Veränderungsmessung hinsichtlich des Therapieerfolgs voneinander.

Insgesamt waren Frauen den Befunden der eingesetzten Instrumente zufolge schwerer erkrankt als Männer. Ebenso waren Patienten mit Persönlichkeitsstörungen stärker beeinträchtigt als jene ohne Persönlichkeitsstörungen. Patienten mit Cluster-B-Persönlichkeitsstörungen wiesen lediglich bei der Messung mit der SCL-90-R einen höheren Gesamtschwereindex auf als Patienten mit Cluster-C-Persönlichkeitsstörungen.

Die Anwendung der Einschlusskriterien führte nicht zu signifikanten Verzerrungen hinsichtlich der Geschlechterverteilung, des Vorliegens von Persönlichkeitsstörungen und hinsichtlich soziodemographischer Parameter mit Ausnahme des Schulabschlusses zwischen der Grundgesamtheit (d. h. aller im Untersuchungszeitraum behandelten Patienten) und der hier analysierten Stichprobe. Der Anteil an Patienten mit Cluster-B- und Cluster-C-Persönlichkeitsstörung unterschied sich allerdings in den Stichproben vor und nach Anwendung der Einschlusskriterien. Bei soziodemographischen Unterschieden zwischen den untersuchten Gruppen bestand kein Anhalt dafür, dass diese die Ergebnisse dieser Arbeit verfälschen. Eine logistische Regression ergab einen statistisch signifikanten, jedoch mit einer Varianzaufklärung von 2,3 % lediglich kleinen Einfluss der soziodemographischen Parameter auf das Auftreten von Erfolg bzw. Misserfolg. Somit ist nicht von einer Verfälschung der Ergebnisse auszugehen.

Insgesamt zeigt diese Arbeit, dass Misserfolge in der stationären Psychotherapie häufig sind. Dabei konnten unter Berücksichtigung sämtlicher Testergebnisse weder das Geschlecht noch das Vorliegen von Persönlichkeitsstörungen oder deren Differenzierung in Cluster-B- und Cluster-C-Persönlichkeitsstörungen als Risikofaktoren für Misserfolge ausgemacht werden. Dieses steht im Einklang mit den Ergebnissen anderer Studien. Im Vergleich zu den meisten dieser Studien lag in dieser Arbeit eine größere Stichprobe vor, welche eine höhere Repräsentativität erlaubt. Außerdem wurde eine umfassende multimodale Diagnostik und Auswertung auf Einzelfallebene angewendet, woraus sich ein Informationsgewinn gegenüber früheren Studien ergibt, die sich auf Effektstärken oder einzelne Erhebungsmethoden beschränken. Über frühere Studien hinausgehend wurden Therapiemisserfolge auch unter Berücksichtigung interpersoneller Aspekte definiert und analysiert und der Einfluss von Persönlichkeitsstörungen auf das Therapieergebnis untersucht. Weitere Stärken, aber auch Limitationen der Studie wurden methodenkritisch im Vergleich zur Literatur diskutiert. Da die Studie die Therapieergebnisse zum Zeitpunkt der Entlassung und nicht zu einem späteren Nacherhebungszeitpunkt untersucht, erlaubt sie beispielsweise keine Vorhersage des Therapieergebnisses zu einem späteren Zeitpunkt, sodass diesbezüglich weiterer Forschungsbedarf besteht. Bei fehlendem Nachweis eines relevanten Einflusses der untersuchten Parameter auf Therapiemisserfolge bleibt der Bedarf bestehen, Einflussfaktoren für Misserfolge, z. B. in prospektiven naturalistischen Studien mittels multimodaler Diagnostik, zu finden.

6 Literaturverzeichnis

Backhaus K, Erichson B, Plinke W, Weiber R: Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. 13. Auflage; Springer, Berlin 2011

Barghaan D, Harfst T, Watzke B, Dirmaier J, Koch U, Schulz H (2007): Merkmale stationärer psychotherapeutischer Versorgung in Deutschland. Psychother Dialog 8, 79-84

Barghaan D, Schulz H, Koch U, Watzke B (2009): Versorgungsstrukturen im stationären Setting in Deutschland: Verteilung von Einzel- und Gruppentherapie und deren psychotherapeutischen Ausrichtungen. Gruppenpsychother Gr 45, 83-103

Bassler M (1995): Prognosefaktoren für den Erfolg von psychoanalytisch fundierter stationärer Psychotherapie. Z Psychosom Med Psychoanal 41, 77-97

Baumann U: Psychodiagnostische Verfahren zur Therapieindikation und Effektkontrolle. In: Bastine R, Fiedler PA, Grawe K, Schmidtchen S, Sommer G (Hrsg.): Grundbegriffe der Psychotherapie. Edition Psychologie, Weinheim 1982, 287-292

Baumann U, Stieglitz RD: Psychodiagnostik psychischer Störungen: Allgemeine Grundlagen. In: Stieglitz RD, Baumann U, Freyberger HJ (Hrsg.): Psychodiagnostik in klinischer Psychologie, Psychiatrie, Psychotherapie. 2. Auflage; Thieme, Stuttgart 2001, 3-20

Baumann U, Sodemann U, Tobien H (1980): Direkte versus indirekte Veränderungsdiagnostik. Z Differentielle Diagnostische Psychol 1, 201-216

Bereiter C: Some Persisting Dilemmas in the Measurement of Change. In: Harris CW (Hrsg.): Problems in Measuring Change. University of Wisconsin Press, Madison 1963, 3-20

Beutel M, Subic-Wrana C: Stationäre integrative Psychotherapie. In: Arolt V, Kersting A (Hrsg.): Psychotherapie in der Psychiatrie: Welche Störung behandelt man wie? Springer, Berlin 2010, 387-403

Beutel M, Doering S, Leichsenring F, Reich G: Psychodynamische Psychotherapie: Störungsorientierung und Manualisierung in der therapeutischen Praxis. Hogrefe, Göttingen 2010

Beutler LE, Crago M: Self-Report Measures of Psychotherapy Outcome. In: Lambert MJ, Christensen ER, De Julio SS (Hrsg.): The assessment of psychotherapy outcome. Wiley, New York 1983, 453-497

Bortz J, Schuster C: Statistik für Human- und Sozialwissenschaftler. 7. Auflage; Springer, Berlin 2010

BPtK (2014): BPtK-Studie zur stationären Versorgung psychisch kranker Menschen: Ergebnisse einer Befragung der in Krankenhäusern angestellten Psychotherapeuten. https://www.bptk.de/wp-content/uploads/2019/01/20140626_BPtK-

Studie_stationaeren_Versorgung_psychisch_kranker_Menschen.pdf; abgerufen am 21.11.2019

Bühl A: SPSS 14: Einführung in die moderne Datenanalyse. 10. Auflage; Pearson Studium, München 2006

Bühl A: SPSS 23: Einführung in die moderne Datenanalyse. 15. Auflage; Pearson, Hallbergmoos 2016

Bühner M: Einführung in die Test- und Fragebogenkonstruktion. 3. Auflage; Pearson Studium, München 2011

Bühner M, Ziegler M: Statistik für Psychologen und Sozialwissenschaftler. Pearson Studium, München 2009

Cohen J: Statistical Power Analysis for the Behavioral Sciences: Second Edition. Lawrence Erlbaum Associates, Hillsdale 1988

Cuijpers P, Li J, Hofmann SG, Andersson G (2010): Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. Clin Psychol Rev 30, 768-778

Davies-Osterkamp S, Strauss B, Schmitz N (1996): Interpersonal Problems as Predictors of Symptom Related Treatment Outcome in Longterm Psychotherapy. Psychother Res 6, 164-176

Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P (2011): Single-Center Trials Show Larger Treatment Effects Than Multicenter Trials: Evidence From a Meta-epidemiologic Study. Ann Intern Med 155, 39-51

Deneke FW: Analytische Gruppentherapie: Eine Prozess- und Erfolgsstudie. Verlag für Medizinische Psychologie im Verlag Vandenhoeck & Ruprecht, Göttingen 1982

DGPPN, DGPM, DKPM, DGP, DGKJP (2009): Persönlichkeitsstörungen: S2-Leitlinie der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN), Deutschen Gesellschaft für Psychosomatische Medizin und Ärztliche Psychotherapie (DGPM), des Deutschen Kollegiums für Psychosomatische Medizin (DKPM), der Deutschen Gesellschaft für Psychologie (DGP), Fachgruppe Klinische Psychologie und Psychotherapie, Deutschen Gesellschaft für Kinder- und Jugendpsychiatrie und -psychotherapie (DGKJP) (S2 Praxisleitlinien in Psychiatrie und Psychotherapie). Band 1; Steinkopff, Darmstadt 2009.

Döring N, Bortz J: Datenerhebung. In: Döring N, Bortz J (Hrsg.): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Auflage; Springer, Berlin 2016a, 321-577

Döring N, Bortz J: Datenanalyse. In: Döring N, Bortz J (Hrsg.): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Auflage; Springer, Berlin 2016b, 597-784

Döring N, Bortz J: Bestimmung von Teststärke, Effektgröße und optimalem Stichprobenumfang. In: Döring N, Bortz J (Hrsg.): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Auflage; Springer, Berlin 2016c, 807-866

Duller C: Einführung in die Statistik mit EXCEL und SPSS: Ein anwendungsorientiertes Lehr- und Arbeitsbuch. 3. Auflage; Springer Gabler, Berlin 2013

Eaton NR, Greene AL (2017): Personality disorders: community prevalence and sociodemographic correlates. Curr Opin Psychol 21, 28-32

Falkai P, Wittchen HU, Döpfner M, Gaebel W, Maier W, Rief W, Saß H, Zaudig M (Hrsg.): Diagnostisches und Statistisches Manual psychischer Störungen: DSM-5 (American Psychiatric Association (APA)). Hogrefe, Göttingen 2015

Field AP: Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll. 4. Auflage; Sage, Los Angeles 2013

Franke GH: SCL-90-R: Symptom-Checkliste von L. R. Derogatis – Deutsche Version – Manual. 2. Auflage; Beltz Test, Göttingen 2002

Franke GH: SCL-90-R: Die Symptom-Checkliste von Derogatis – Deutsche Version. In: Brähler E, Schumacher J, Strauß B (Hrsg.): Diagnostische Verfahren in der Psychotherapie. 2. Auflage; Hogrefe, Göttingen 2003, 299-303

Franke GH: SCL-90-S: Symptom-Checklist-90®-Standard. Manual. Hogrefe, Göttingen 2014

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

Fromm S: Binäre logistische Regressionsanalyse: Eine Einführung für Sozialwissenschaftler mit SPSS für Windows (Bamberger Beiträge zur empirischen Sozialforschung). Universität Bamberg, Bamberg 2005.

Fydrich T: Persönlichkeitsstörungen. In: Röhrle B, Caspar F, Schlottke PF (Hrsg.): Lehrbuch der klinisch-psychologischen Diagnostik. Kohlhammer, Stuttgart 2008, 755-769

Geiser F, Imbierowicz K, Conrad R, Schilling G, Liedtke R (2001): Zur Unterscheidung von „geheilten“ und „gebesserten“ sowie von „unveränderten“ und „verschlechterten“ Patienten in einer Therapieerfolgsstudie. Z Psychosom Med Psychother 47, 250-261

Geyer M, Plöttner G, Machold C, Dann K, Schmidt B: Lässt sich ein Psychotherapieerfolg schon während der stationären Behandlung erkennen? In: Geyer M, Plöttner G, Villmann

T (Hrsg.): Psychotherapeutische Reflexionen gesellschaftlichen Wandels. VAS - Verlag für Akademische Schriften, Frankfurt am Main 2003, 293-304

Goldhammer F, Hartig J: Interpretation von Testresultaten und Testeichung. In: Moosbrugger H, Kelava A (Hrsg.): Testtheorie und Fragebogenkonstruktion. 2. Auflage; Springer, Berlin 2012, 173-201

Grawe K: Der Veränderungsprozeßbogen (VPB). In: Zielke M (Hrsg.): Diagnostik in der Psychotherapie. Kohlhammer, Stuttgart 1982, 231-252

Haase M, Frommer J, Franke GH, Hoffmann T, Schulze-Muetzel J, Jäger S, Grabe HJ, Spitzer C, Schmitz N (2008): From symptom relief to interpersonal change: Treatment outcome and effectiveness in inpatient psychotherapy. Psychother Res 18, 615-624

Habermeyer V, Herpertz SC (2009): Diagnose und Therapie von Persönlichkeitsstörungen. Fortschr Neurol Psychiatr 77, 532-546

Hartkamp N, Hildenbrand G: Stationäre psychosomatisch-psychotherapeutische Behandlung. In: Janssen PL, Joraschky P, Tress W (Hrsg.): Leitfaden Psychosomatische Medizin und Psychotherapie: Orientiert an den Weiterbildungsrichtlinien der Bundesärztekammer. 2. Auflage; Deutscher Ärzteverlag, Köln 2009, 663-672

Haupt ML, Linden M, Strauß B: Definition und Klassifikation von Psychotherapie-Nebenwirkungen. In: Linden M, Strauß B (Hrsg.): Risiken und Nebenwirkungen von Psychotherapie: Erfassung, Bewältigung, Risikovermeidung. Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin 2013, 1-13

Heigl-Evers A, Heigl FS, Ott J: Zur Theorie und Praxis der psychoanalytisch-interaktionellen Gruppentherapie. In: Heigl-Evers A, Ott J (Hrsg.): Die psychoanalytisch-interaktionelle Methode: Theorie und Praxis. Vandenhoeck & Ruprecht, Göttingen 1994, 226-264

Heigl-Evers A, Heigl FS, Ott J: Abriß der Psychoanalyse und der analytischen Psychotherapie. In: Heigl-Evers A, Heigl F, Ott J, Rüger U (Hrsg.): Lehrbuch der Psychotherapie. 3. Auflage; Gustav Fischer, Lübeck 1997, 3-270

Hoffmann SO, Rudolf G, Strauß B (2008): Unerwünschte und schädliche Wirkungen von Psychotherapie: Eine Übersicht mit dem Entwurf eines eigenen Modells. Psychotherapeut 53, 4-16

Horowitz LM, Strauß B, Kordy H: IIP-C. Beltz Test, Göttingen 1994

Horowitz LM, Strauß B, Kordy H: IIP-D: Inventar zur Erfassung Interpersonaler Probleme – Deutsche Version – Manual. 2. Auflage; Beltz Test, Göttingen 2000

Huber D, Henrich G, Klug G (2007): The Inventory of Interpersonal Problems (IIP): Sensitivity to change. Psychother Res 17, 474-481

Jacobi F: Misserfolgsforschung in der Verhaltenstherapie. In: Dohrenbusch R, Kaspers F (Hrsg.): Fortschritte der Klinischen Psychologie und Verhaltensmedizin. Pabst Science Publishers, Lengerich 2001, 323-346

Jacobson NS, Truax P (1991): Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59, 12-19

Junge A, Ahrens S (1996): Stationäre psychosomatische Behandlung: Patientenmerkmale und Behandlungserfolg. Psychother Psychosom Med Psychol 46, 430-437

Köck K, Lutz W (2012): Therapieverlaufs- und Ergebnismessung im Rahmen einer patientenorientierten Psychotherapieforschung. Psychother Dialog 13, 10-15

Konzag TA, Bandemer-Greulich U, Bahrke U, Fikentscher E (2004): Therapeutische Beziehung und Therapieerfolg bei der stationären Psychotherapie von Persönlichkeitsstörungen. Z Psychosom Med Psychother 50, 394-405

Koo TK, Li MY (2016): A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 15, 155-163

Kordy H, Hannöver W, Bauer S: Das Stuttgart-Heidelberger Modell zur Qualitätssicherung in der stationären Psychotherapie. In: Härter M, Linster HW, Stieglitz RD (Hrsg.): Qualitätsmanagement in der Psychotherapie: Grundlagen, Methoden und Anwendung. Hogrefe, Göttingen 2003, 289-304

Krampen G, von Delius A (1981): Zur direkten Messung subjektiv erlebter gesundheitlicher Veränderungen. Med Psychol 7, 166-174

Krampen G, Hank P: Prozessdiagnostik und kontrollierte Praxis. In: Röhrle B, Caspar F, Schlottke PF (Hrsg.): Lehrbuch der klinisch-psychologischen Diagnostik. Kohlhammer, Stuttgart 2008, 300-329

Kriebel R, Paar GH, Schmitz-Buhl SM, Raatz U (2001): Veränderungsmessung mit dem Veränderungsfragebogen (VEV): Entwicklung einer Kurzform und deren Anwendung in der Psychosomatischen Rehabilitation. Prax Klin Verhaltensmed Rehabil, 20-32

Kuckartz U, Rädiker S, Ebert T, Schehl J: Statistik: Eine verständliche Einführung. 2. Auflage; Springer, Wiesbaden 2013

Leichsenring F, Rüger U (2004): Psychotherapeutische Behandlungsverfahren auf dem Prüfstand der Evidence Based Medicine (EBM). Z Psychosom Med Psychother 50, 203-217

Liebherz S, Rabung S (2013): Wirksamkeit psychotherapeutischer Krankenhausbehandlung im deutschsprachigen Raum: Eine Meta-Analyse. Psychother Psychosom Med Psychol 63, 355-364

Liebler A, Biella M, Schulz W, Hartmann U (2004): Differenzielle Therapieverläufe und Behandlungserfolg in der stationären Gruppenpsychotherapie. Gruppenpsychother Gr 40, 1-21

Lienert GA, Raatz U: Testaufbau und Testanalyse. 6. Auflage; Beltz Psychologie Verlags Union, Weinheim 1998

Lilienfeld SO (2007): Psychological Treatments That Cause Harm. Perspect Psychol Sci 2, 53-70

Linden M (2013): How to Define, Find and Classify Side Effects in Psychotherapy: From Unwanted Events to Adverse Treatment Reactions. Clin Psychol Psychother 20, 286-296

Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channabasavanna SMW, Coid B, Dahl A, Diekstra RF, Ferquson B et al. (1994): The International Personality Disorder Examination: The World Health Organization/Alcohol, Drug Abuse, and Mental Health Administration International Pilot Study of Personality Disorders. Arch Gen Psychiatry 51, 215-224

Maier W, Albus M, Bech P: Validitätskriterien für die Veränderungsmessung mit Ratingskalen. In: Baumann U (Hrsg.): Veränderungsmessung in Psychiatrie und klinischer Psychologie: Theoretische, methodische und empirische Beiträge. Profil, München 1990, 286-306

Maier W, Lichtermann D, Klingler T, Heun R, Hallmayer J (1992): Prevalences of Personality Disorders (DSM-III-R) in the Community. J Pers Disord 6, 187-196

Martens U, Rempel I, Zipfel S, Enck P, Teufel M (2014): Behandlungsabbrüche während einer stationären Psychotherapie. Z Psychosom Med Psychother 60, 238-250

Mirowsky J, Ross CE (1995): Sex differences in distress: real or artifact? Am Sociol Rev 60, 449-468

Mohr DC (1995): Negative Outcome in Psychotherapy: A Critical Review. Clin Psychol (New York) 2, 1-27

Möller HJ: Möglichkeiten und Grenzen von Selbstbeurteilungsskalen zur Verlaufsbeurteilung depressiver Symptomatik im Rahmen der Therapieevaluation. In: Baumann U (Hrsg.): Veränderungsmessung in Psychiatrie und klinischer Psychologie: Theoretische, methodische und empirische Beiträge. Profil, München 1990, 307-328

Möller-Leimkühler AM (2005): Geschlechtsrolle und psychische Erkrankung. J Neurol Neurochir Psychiatr 6, 29-35

Mößner M, Kordy H (2014): Veränderungsmessung und Ergebnismonitoring in der klinischen Routine. Z Psychiatr Psych Ps 62, 93-100

Nübling R, Steffanowski A, Wittmann WW, Schmidt J (2004): Strategien der Ergebnismessung am Beispiel der psychosomatischen Rehabilitation. Prax Klin Verhaltensmed Rehabil, 35-44

Ogles BM: Measuring change in psychotherapy research. In: Lambert MJ (Hrsg.): Bergin and Garfield's handbook of psychotherapy and behavior change. 6. Auflage; John Wiley & Sons, Hoboken 2013, 134-166

Oster J, Müller G, von Wietersheim J (2009): "Wer profitiert?"- Patientenmerkmale als Erfolgsprädiktoren in der psychosomatischen Rehabilitation. Rehabilitation (Stuttg) 48, 95-102

Paar GH (1997): Das Qualitätssicherungsprogramm der Rentenversicherungsträger: Perspektive des Anwenders aus einer psychosomatischen Fachklinik. Psychotherapeut 42, 156-162

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 Klin Psychol Psychother (Gott) 33, 165-171

Reimer, C., Rüger, U. (Hrsg.) (2012): Psychodynamische Psychotherapien: Lehrbuch der tiefenpsychologisch fundierten Psychotherapieverfahren. 4. Auflage; Springer, Berlin 2012

Reuter L, Scheidt CE (2014): Vorzeitige Beendigung der Therapie in der stationären psychosomatischen Krankenhausbehandlung und Rehabilitation. Psychother Psychosom Med Psychol 64, 297-308

Reuter L, Bengel J, Scheidt CE (2014): Therapie-Non-Response in der psychosomatischen Krankenhausbehandlung und Rehabilitation – Eine systematische Übersicht. Z Psychosom Med Psychother 60, 121-145

Riecher-Rössler A (2000): Psychische Erkrankungen bei Frauen – einige Argumente für eine geschlechtersensible Psychiatrie und Psychotherapie. Z Psychosom Med Psychother 46, 129-139

Rief W, Stock C, Geissner E, Fichter MM (1994): Wenn Patient und Therapeut unterschiedlicher Meinung sind – Diskrepanzen in Veränderungsbeurteilungen. Psychother Psychosom Med Psychol 44, 235-239

Rost J: Lehrbuch Testtheorie – Testkonstruktion. 2. Auflage; Verlag Hans Huber, Bern 2004

Rudolf G: Strukturbezogene Psychotherapie: Leitfaden zur psychodynamischen Therapie struktureller Störungen. 3. Auflage; Schattauer, Stuttgart 2013

Rudolf G, Rüger U (2001): Zur Differentialindikation zwischen tiefenpsychologisch fundierter und analytischer Psychotherapie. Psychotherapeut 46, 216-219

Rudolf G, Rüger U: Psychodynamische Therapieverfahren: Welche nichtpsychodynamischen Ansätze lassen sich integrieren? In: Reimer C, Rüger U (Hrsg.): Psychodynamische Psychotherapien: Lehrbuch der tiefenpsychologisch fundierten Psychotherapieverfahren. 4. Auflage; Springer, Berlin 2012, 25-33

Rudolf G, Jakobsen T, Micka R, Schumann E (2004): Störungsbezogene Ergebnisse psychodynamisch-stationärer Psychotherapie. Z Psychosom Med Psychother 50, 37-52

Rüger U, Reimer C: Gemeinsame Merkmale und Charakteristika psychodynamischer Psychotherapieverfahren. In: Reimer C, Rüger U (Hrsg.): Psychodynamische Psychotherapien: Lehrbuch der tiefenpsychologisch fundierten Psychotherapieverfahren. 4. Auflage; Springer, Berlin 2012, 3-24

Sammet I, Staats H, Schauenburg H (2004): Beziehungserleben und Therapieergebnis in stationärer Psychotherapie. Z Psychosom Med Psychother 50, 376-393

Schauenburg H, Sammet I, Strack M (2001): Verlauf der Symptombelastung und Vorhersage des Behandlungserfolges in der stationären Psychotherapie. Z Psychosom Med Psychother 47, 380-395

Schellberg D: Reliabilität des BSS. In: Schepank H (Hrsg.): Der Beeinträchtigungs-Schwere-Score (BSS): Ein Instrument zur Bestimmung der Schwere einer psychogenen Erkrankung: Handanweisung. Beltz Test, Göttingen 1995, 18-20

Schepank H (Hrsg.) (1995): Der Beeinträchtigungs-Schwere-Score (BSS): Ein Instrument zur Bestimmung der Schwere einer psychogenen Erkrankung: Handanweisung. Beltz Test, Göttingen 1995

Schepank H: BSS: Der Beeinträchtigungs-Schwere-Score. In: Brähler E, Schumacher J, Strauß B (Hrsg.): Diagnostische Verfahren in der Psychotherapie. 2. Auflage; Hogrefe, Göttingen 2003, 68-72

Schmidt-Atzert L, Amelang M: Psychologische Diagnostik. 5. Auflage; Springer, Berlin 2012

Schulte D: Optimierung der Entscheidungsprozesse von Psychotherapeuten. In: Freyberger HJ, Heuft G, Ziegenhagen DJ (Hrsg.): Ambulante Psychotherapie: Transparenz, Effizienz, Qualitätssicherung. Schattauer, Stuttgart 2000, 15-41

Schulz H, Barghaan D, Harfst T, Koch U: Gesundheitsberichterstattung des Bundes Heft 41: Psychotherapeutische Versorgung. Robert Koch-Institut, Berlin 2008.

Schulz H, Barghaan D, Koch U, Harfst T: Die Versorgung von Patienten mit psychischen Störungen. In: Wittchen HU, Hoyer J (Hrsg.): Klinische Psychologie & Psychotherapie. 2. Auflage; Springer, Berlin 2011, 361-380

Seidenstücker G, Baumann U: Multimethodale Diagnostik. In: Baumann U, Berbalk H, Seidenstücker G (Hrsg.): Klinische Psychologie: Trends in Forschung und Praxis. Band 1; Huber, Bern 1978, 134-182

Shadish WR, Navarro AM, Matt GE, Phillips G (2000): The Effects of Psychological Therapies Under Clinically Representative Conditions: A Meta-Analysis. Psychol Bull 126, 512-529

Shrout PE, Fleiss JL (1979): Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull 86, 420-428

Sozialgesetzbuch (SGB V), Fünftes Buch, Gesetzliche Krankenversicherung vom 20.12.1988, zuletzt geändert am 09.08.2019

Spitzer C, Masuhr O, Jaeger U, Brünig S, Dümpelmann M, Dally A (2012): Misserfolge in der stationären Therapie von Patienten mit Persönlichkeitsstörungen. Psychotherapeut 57, 417-424

Spitzer C, Masuhr O, Jaeger U, Brünig S, Dally A, Fricke-Neef C, Dinger U, Zimmermann J (2015): Rolle der Bezugstherapeuten in der stationären Psychotherapie: Klinische Beobachtungsstudie zur Bedeutung für den symptombezogenen Behandlungserfolg. Psychotherapeut 60, 302-309

Spitzer C, Rullkötter N, Dally A (2016): Stationäre Psychotherapie. Nervenarzt 87, 99-108

Spitzer C, Masuhr O, Jaeger U, Euler S (2018): Pseudogesunde: Herausforderung für psychotherapeutische Versorgung und empirische Psychotherapieforschung. Psychotherapeut 63, 483-490

Spitzer C, Masuhr O, Jaeger U, Euler S (2019): Pseudogesunde Patienten in der stationären Psychotherapie – Ein Annäherungsversuch. Psychother Psychosom Med Psychol 69, 445-452

Statistisches Bundesamt (Destatis): Diagnosedaten der Patienten und Patientinnen in Krankenhäusern (einschl. Sterbe- und Stundenfälle): 2015 (Fachserie 12 Reihe 6.2.1). Statistisches Bundesamt, Wiesbaden 2016a.

Statistisches Bundesamt (Destatis): Bevölkerungsfortschreibung auf Grundlage des Zensus 2011: 2015 (Fachserie 1 Reihe 1.3). Statistisches Bundesamt, Wiesbaden 2016b.

Statistisches Bundesamt (Destatis): Grunddaten der Krankenhäuser: 2016 (Fachserie 12 Reihe 6.1.1). Statistisches Bundesamt, Wiesbaden 2017a.

Statistisches Bundesamt (Destatis): Krankheitskosten: 2015 (Fachserie 12 Reihe 7.2.1). Statistisches Bundesamt, Wiesbaden 2017b.

Stieglitz RD: Selbst- und Fremdbeurteilung in der psychologisch-psychiatrischen Diagnostik und Therapieforschung. In: Janssen PL, Schneider W (Hrsg.): Diagnostik in Psychotherapie und Psychosomatik. Gustav Fischer, Stuttgart 1994, 37-64

Stieglitz RD, Baumann U: Veränderungsmessung. In: Stieglitz RD, Baumann U, Freyberger HJ (Hrsg.): Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie. 2. Auflage; Thieme, Stuttgart 2001, 21-38

Stieglitz RD, Freyberger HJ: Selbstbeurteilungsverfahren. In: Stieglitz RD, Baumann U, Freyberger HJ (Hrsg.): Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie. 2. Auflage; Thieme, Stuttgart 2001, 83-94

Stieglitz RD, Hiller W (2014): Strategien und Instrumente der Veränderungsmessung. Z Psychiatr Psych Ps 62, 101-111

Stieglitz RD, Baumann U, Tobien H, von Zerssen D (1980): Zur Stichproben- und Zeitinvarianz von Testkennwerten bei einer Beschwerdenliste. Z Exp Angew Psychol 27, 631-654

Stieglitz RD, Ahrens B, Freyberger HJ: Fremdbeurteilungsverfahren. In: Stieglitz RD, Baumann U, Freyberger HJ (Hrsg.): Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie. 2. Auflage; Thieme, Stuttgart 2001, 95-106

Strauß B: IIP-D: Inventar zur Erfassung interpersonaler Probleme. In: Brähler E, Schumacher J, Strauß B (Hrsg.): Diagnostische Verfahren in der Psychotherapie. 2. Auflage; Hogrefe, Göttingen 2003, 206-210

Streeck U, Leichsenring F: Handbuch psychoanalytisch-interaktionelle Therapie: Behandlung von Patienten mit strukturellen Störungen und schweren Persönlichkeitsstörungen. 2. Auflage; Vandenhoeck & Ruprecht, Göttingen 2011

Strupp HH (1980): Success and Failure in Time-Limited Psychotherapy: A Systematic Comparison of Two Cases: Comparison 1. Arch Gen Psychiatry 37, 595-603

Strupp HH, Hadley SW, Gomes-Schwartz B: Psychotherapy for better or worse: The problem of negative effects. J. Aronson, New York 1977

S2-Leitlinie Persönlichkeitsstörungen: siehe DGPPN, DGPM, DKPM, DGP, DGKJP 2009

Universitätsklinikum Heidelberg - Forschungsstelle für Psychotherapie (2015): Errata zum Manual der 2. Auflage des IIP-D. https://www.klinikum.uni-heidelberg.de/IIP-Errata.143497.0.html; abgerufen am: 10.03.2019

Waldmann HC, Petermann F (2014): Veränderungsmessung – Methodische Vorschläge für Forschung und klinische Praxis. Z Psychiatr Psych Ps 62, 85-92

Wissenschaftlicher Beirat Psychotherapie (2005): Stellungnahme zur Psychodynamischen Psychotherapie bei Erwachsenen. Dtsch Arztebl 102, A 73-75

Wolf M, Gallas C, Kordy H (2005): Zur diskriminativen Validität des Inventars zur Erfassung Interpersonaler Probleme (IIP): Interpersonale Probleme von Patienten mit Persönlichkeitsstörungen. Z Klin Psychol Psychiatr Psychother 34, 233-240

Wolfersdorf M (2008): Suizidalität. Nervenarzt 79, 1319-1334

Zimmer B, Moessner M (2012): Therapieevaluation in der stationären Psychotherapie mit Web-AKQUASI. Psychother Dialog 13, 68-72

Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. med. C. Spitzer für die Möglichkeit, diese Dissertation zu erstellen, und für die allzeit exzellente wissenschaftliche Begleitung. Für seine geduldige und wohlwollende Unterstützung und Betreuung, insbesondere für die Bereitstellung von Daten und Materialien, konkrete Hilfestellungen, wertvolle konstruktive Vorschläge und ein ausführliches Feedback, möchte ich mich ganz herzlich bedanken.

Des Weiteren danke ich Herrn Dipl.-Psych. O. Masuhr für die kompetente Beratung bei Fragestellungen zur Datenauswertung und Statistik, die ein eigenständiges Bearbeiten des Themas ermöglicht hat, und für seine zielorientierten Rückmeldungen, die zur Fertigstellung dieser Dissertation beigetragen haben.