

**Vorhersage des längerfristigen Behandlungserfolgs
anhand von Symptomverläufen und der Dauer
stationärer Psychotherapie**

Dissertation

zur Erlangung der Würde des Doktors der Philosophie

am Fachbereich Psychologie

Fakultät für Erziehungswissenschaft, Psychologie und Bewegungswissenschaft
der Universität Hamburg

vorgelegt von
Hanne Melchior

Hamburg 2011

Erster Dissertationsgutachter:	Prof. Dr. Dr. Uwe Koch-Gromus
Zweiter Dissertationsgutachter:	Prof. Dr. Bernhard Dahme
Erster Disputationsgutachter:	Prof. Dr. Holger Schulz
Zweiter Disputationsgutachter:	PD Dr. Andreas von Leupoldt
Disputationsvorsitzende:	Prof. Dr. Monika Bullinger
Tag der Disputation:	07. November 2011

Die vorliegende Arbeit wäre ohne die Unterstützung, die ich von verschiedenen Seiten erhalten habe, nicht in dieser Form möglich gewesen.

Mein besonderer Dank geht an PD Dr. Birgit Watzke, Prof. Dr. Holger Schulz und Prof. Dr. Dr. Uwe Koch-Gromus, die die Rahmenbedingungen für die Erstellung dieser Arbeit geschaffen haben und diese in allen Phasen ihrer Entstehung sehr fördernd begleitet haben. Bedanken möchte ich mich auch bei Kerstin Hofreuter-Gätgens und Anika Hergert sowie dem gesamten Leitungsteam der „Nachwuchsgruppe Rehabilitationsforschung“ (Prof. Dr. Dr. Uwe Koch-Gromus, Prof. Dr. Holger Schulz, PD Dr. Birgit Watzke, PD Dr. Corinna Bergelt und Prof. Dr. Matthias Morfeld) für die vielen wertvollen Diskussionen und den regen Austausch. Der Willhelmstiftung für Rehabilitationsforschung danke ich für die finanzielle Unterstützung und Ermöglichung der Studie im Rahmen der „Nachwuchsgruppe Rehabilitationsforschung“. Für ihre Kooperation bei der aufwändigen Datenerhebung bedanke ich mich insbesondere bei Sabine Zahn, Prof. Dr. Volker Kollenbaum, Oliver Bohlen, Kurt Langner, Dr. Silke Kleinschmidt und Kristine Klapheck, die die Studie begleitet und durch ihr Engagement in den Kliniken vorangetrieben haben.

Prof. Dr. Dr. Martin Härter gilt mein Dank für die Unterstützung und Förderung meiner Arbeit am Institut für Medizinische Psychologie. Ebenso danke ich allen Kolleginnen und Kollegen, die mir mit ihrem Rat zur Seite standen. Für die vielen wertvollen methodischen Diskussionen und Anregungen, die oft weit über die Methoden hinaus gingen, bedanke ich mich bei Dr. Levente Kriston.

Für die Unterstützung in der gesamten Zeit der Entstehung der Arbeit, für die Geduld und die willkommenen Ablenkungen möchte ich mich bei meinen Freunden bedanken. Vor allem möchte ich mich bei meiner Familie bedanken, die mich immer und uneingeschränkt darin unterstützt hat, meine Ziele zu verfolgen und mir bei der Arbeit den Rückhalt gab, den ich brauchte.

Inhaltsverzeichnis

Zusammenfassung	1
1 Einleitung	3
2 Theoretischer Hintergrund	7
2.1 Psychotherapeutische Versorgungsstruktur in Deutschland.....	7
2.1.1 Behandlungsstrukturen.....	7
2.1.2 Behandlungsprozesse.....	9
2.2 Psychotherapieforschung und Versorgungsforschung.....	12
2.3 Wirksamkeit stationärer psychotherapeutischer Behandlungen.....	15
2.3.1 Maße und Bewertung der Wirksamkeit stationärer Psychotherapie.....	15
2.3.2 Empirische Befunde zur Wirksamkeit stationärer Psychotherapie.....	17
2.4 Vorhersage der Wirksamkeit stationärer Psychotherapie.....	21
2.4.1 Patientenmerkmale als Prädiktoren des Behandlungserfolgs.....	21
2.4.2 Prozessmerkmale als Prädiktoren des Behandlungserfolgs.....	25
2.5 Behandlungsdauer und Symptomverlauf während der Behandlung als Prädiktoren des Behandlungserfolgs.....	27
2.5.1 Dosis und Dauer.....	28
2.5.2 Theoretischer Hintergrund zum Dosis-/Dauer-Wirkungs-Zusammenhang.....	29
2.5.3 Empirische Befunde zum Dosis-Wirkungs-Zusammenhang und Symptomverlauf während der Behandlung.....	31
2.5.4 Das Konzept der Early Response.....	35
2.5.5 Early Response als Prädiktor des Behandlungserfolgs.....	38
2.5.6 Inhaltliche Erklärungsansätze der Early Response.....	40
2.5.7 Neuere Methoden der Untersuchung von Symptomverlaufgruppen während der Behandlung.....	40
2.6 Vorhersage der Behandlungsdauer und des Behandlungsverlaufs.....	43
2.6.1 Vorhersage der Behandlungsdauer.....	44
2.6.2 Vorhersage des Symptomverlaufs während der Behandlung.....	45
3 Fragestellungen und Hypothesen	49
3.1 Zielsetzung der Untersuchung.....	49
3.2 Fragestellungen.....	50

3.3	Hypothesen	54
4	Methode.....	59
4.1	Studiendesign.....	59
4.2	Durchführung der Datenerhebung.....	60
4.2.1	Setting.....	60
4.2.2	Datenerhebung und Datenschutz	61
4.2.3	Ein- und Ausschlusskriterien	63
4.3	Operationalisierung der zu untersuchenden Merkmale.....	64
4.3.1	Operationalisierung des primären Outcomes und des Verlaufs	64
4.3.2	Operationalisierung des sekundären Outcomes.....	65
4.3.3	Operationalisierung der Prädiktormerkmale	66
4.3.4	Operationalisierung der theoriegeleiteten Verlaufsgruppen	70
4.4	Datenmanagement.....	72
4.4.1	Aufbereitung der Daten.....	72
4.4.2	Umgang mit fehlenden Werten	73
4.5	Statistische Auswertungsmodelle.....	75
4.5.1	Stichprobenbeschreibung und Drop-Out-Analysen	75
4.5.2	Verlaufsanalyse	76
4.5.3	Prädiktoren- und Outcome-Analysen.....	80
4.5.4	Bestimmung und Interpretation statistischer und klinischer Effekte	83
4.6	Poweranalysen.....	86
4.6.1	Statistische Poweranalyse.....	86
4.6.2	Stichprobenkalkulation und Erhebungszeitraum	87
5	Ergebnisse	89
5.1	Stichproben, Repräsentativitäts- und Drop-Out-Analysen.....	89
5.1.1	Stichprobenbildung	89
5.1.2	Repräsentativitätsanalysen.....	90
5.1.3	Stichprobenbeschreibung und Drop-Out-Analysen	92
5.2	Ergebnisse der explorativen Verlaufsanalyse (Fragestellung 1 A).....	107
5.2.1	Identifikation der Verlaufsklassen	107
5.2.2	Beschreibung der Verlaufsklassen	111
5.2.3	Inhaltliche Modifikation der Verlaufsklassen.....	117
5.3	Ergebnisse der theoriegeleiteten Verlaufsanalyse (Fragestellung 1 B).....	119

5.4	Vergleich der explorativen und der theoriegeleiteten Verlaufsguppen (Fragestellung 1 C).....	122
5.5	Wöchentliche Verbesserungsdaten (Fragestellung 1 D).....	125
5.6	Vorhersage der Verlaufsguppen (Fragestellung 2)	129
5.6.1	Vergleiche der explorativen Verlaufsguppen.....	130
5.6.2	Vergleiche der theoriegeleiteten Verlaufsguppen.....	136
5.6.3	Auswahl der Prädiktorvariablen	140
5.6.4	Vorhersage der Verlaufsguppen.....	145
5.7	Vorhersage des primären Outcomes durch die explorativen Verlaufsguppen (Fragestellung 3 A).....	155
5.7.1	Unterscheiden sich die explorativen Verlaufsguppen im primären Outcome?	155
5.7.2	Auswahl der zu kontrollierenden Prädiktoren	158
5.7.3	Vorhersage des primären Outcomes durch die explorativen Verlaufsguppen	160
5.8	Vorhersage des primären Outcomes durch die theoriegeleiteten Verlaufsguppen (Fragestellung 3 B).....	163
5.8.1	Unterscheiden sich die theoriegeleiteten Verlaufsguppen im primären Outcome?	163
5.8.2	Auswahl der zu kontrollierenden Prädiktorvariablen.....	166
5.8.3	Vorhersage des primären Outcomes durch die theoriegeleiteten Verlaufsguppen	166
5.9	Vorhersage des sekundären Outcomes durch die explorativen Verlaufsguppen (Fragestellung 3 C)	169
5.9.1	Unterscheiden sich die explorativen Verlaufsguppen im sekundären Outcome?	169
5.9.2	Auswahl der zu kontrollierenden Prädiktorvariablen.....	173
5.9.3	Vorhersage des sekundären Outcomes durch die explorativen Verlaufsguppen	173
5.10	Vorhersage des sekundären Outcomes durch die theoriegeleiteten Verlaufsguppen (Fragestellung 3 D)	176
5.10.1	Unterscheiden sich die theoriegeleiteten Verlaufsguppen im sekundären Outcome?	176
5.10.2	Auswahl der zu kontrollierenden Prädiktorvariablen.....	179
5.10.3	Vorhersage des sekundären Outcomes durch die theoriegeleiteten Verlaufsguppen	179
5.11	Welche Verlaufsguppen erlauben eine bessere Vorhersage des katamnestic Outcomes?	181

6	Diskussion	185
6.1	Generalisierbarkeit der Ergebnisse	185
6.1.1	Studienteilnahme	185
6.1.2	Drop-Out	188
6.2	Diskussion der Ergebnisse der Fragestellungen und Hypothesen	191
6.2.1	Welche Behandlungsverläufe der Symptombelastung lassen sich identifizieren?	191
6.2.2	Welche Merkmale erlauben eine Vorhersage des Behandlungsverlaufs?	195
6.2.3	Lässt sich das längerfristige Behandlungsergebnis durch den Symptomverlauf während der Behandlung vorhersagen?	208
6.3	Methodische Diskussion	224
6.3.1	Alpha-Fehler und Teststärke	224
6.3.2	Studiendesign, interne und externe Validität	225
6.3.3	Kritische Auseinandersetzung mit der Methode des Growth Mixture Modeling	227
6.3.4	Validität der Prädiktormasse	230
6.3.5	Angemessenheit der Erhebungszeitpunkte	232
6.3.6	Angemessenheit der Verlaufs- und Outcome-Masse	234
6.4	Ausblick	235
7	Literaturverzeichnis	241
	Abbildungsverzeichnis	259
	Tabellenverzeichnis	261
	Anhang	267
	Anhang A: Zusammenhangsmasse der untersuchten Merkmale	269
	Anhang B: Stichprobenbeschreibungen und Drop-Out-Analysen	275
	Anhang C: Vergleiche der explorativen Verlaufgruppen	279
	Anhang D: Vergleiche der theoriegeleiteten Verlaufgruppen	289
	Anhang E: Patienteninformation, Einverständniserklärung, Katamnese-Anschreiben	299

Zusammenfassung

Die stationäre Behandlung psychisch erkrankter Patienten¹ wurde in den letzten Jahren stetig verkürzt, wobei die empirische Basis zum Verlauf der Symptomatik während der Behandlung sowie zur Bestimmung optimaler Behandlungsdauern gering ist. In der vorliegenden Arbeit wurde der Zusammenhang zwischen dem Verlauf der Symptombelastung während der stationären Behandlung und dem Therapieergebnis der Behandlung psychischer/psychosomatischer Erkrankungen untersucht. Dies erfolgte anhand der folgenden drei Fragestellungen: 1. Lassen sich Patienten(-gruppen) mit unterschiedlichen Verläufen der Symptombelastung während der Behandlung identifizieren? 2. Können die identifizierten Symptomverläufe durch bestimmte Patientencharakteristika vorhergesagt werden? 3. Kann das längerfristige Therapieergebnis durch die identifizierten Symptomverläufe vorhergesagt werden? Ziel war es, differenzierte Aussagen über den gesamten Verlauf der stationären Behandlung, über den Zeitpunkt und die Stabilität von Verbesserungen, über die Zuordnung der Patienten zu bestimmten prototypischen Symptomverlaufsgruppen und über den Vorhersagewert des Symptomverlaufs für das längerfristige Behandlungsergebnis zu treffen.

Mittels einer prospektiven multizentrischen Untersuchung im Ein-Gruppen-Studiendesign wurden neben den Erhebungen zu Beginn und bei Beendigung der stationären Behandlung kontinuierliche wöchentliche Messungen während der gesamten stationären Behandlung sowie eine katamnestische Erhebung 6 Monate nach Beendigung der Behandlung durchgeführt. Verlaufsmaß und primäres Outcome stellten dabei die Symptombelastung dar. Als sekundäres Outcome wurden die psychischen Aspekte der gesundheitsbezogenen Lebensqualität erhoben.

In drei Fachkliniken wurde eine konsekutive Stichprobe von 576 Patienten mit psychischen Erkrankungen untersucht, zur Katamnese stand noch eine Stichprobe von 424 Patienten zur Verfügung. Mittels latenter Wachstumskurvenanalysen (Growth Mixture Modeling) konnten explorativ fünf distinkte Verlaufsgruppen identifiziert werden. Zudem wurde eine Verlaufsgruppeneinteilung mit vier Gruppen nach theoriegeleiteten Kriterien vorgenommen. Die Symptomverläufe beider Gruppeneinteilungen ließen sich zu einem bedeutenden Anteil durch Patientenmerkmale vorhersagen. Dabei standen neben einem niedrigen Schulabschluss und der Nichterwerbstätigkeit vor allem Indikatoren für chronifizierte Erkrankungen im Zusammenhang mit einem ungünstigen Verlauf während der Behandlung.

¹ Aus Gründen der besseren Lesbarkeit wird bei Sammelbezeichnungen (*Patienten, Therapeuten, usw.*) durchgängig die grammatikalisch männliche Form benutzt, wobei jeweils männliche und weibliche Personen gleichermaßen gemeint sind.

Ein frühes Ansprechen während der Behandlung zeigt sich bei Patienten mit hohem initialen Leidensdruck und tendenziell geringer Chronifizierung. Hinsichtlich der Varianz im primären und sekundären Outcome konnte, bei statistischer Berücksichtigung fundierter Prädiktoren des Outcomes, noch ein bedeutender Anteil durch die Symptomverläufe erklärt werden. Dabei zeigen vor allem die explorativ ermittelten Verlaufgruppen einen guten Vorhersagewert für das längerfristige Behandlungsergebnis.

Die gewonnenen Ergebnisse werden auf inhaltlicher Ebene diskutiert und in Bezug zu bisherigen empirischen Befunden gesetzt. Ein zweiter Fokus der Diskussion liegt in der methodischen Auseinandersetzung mit den verwendeten Verfahren zur Gruppenbildung, dem Studiendesign und der Operationalisierung von Prädiktor- und Outcome-Maßen. Aus den Ergebnissen können sowohl forschungsrelevante als auch klinische Implikationen zur Untersuchung und Optimierung von Behandlungsprozessen abgeleitet werden. Erkenntnisse zum Verlauf der Symptombelastung während der stationären Behandlung leisten einen wertvollen Beitrag zur Vorhersage des Behandlungsergebnisses und somit zur Evidenzbasierung von Entscheidungen über Behandlungsprozesse.

1 Einleitung

Psychische Erkrankungen zeigen eine weite Verbreitung in der Bevölkerung der westlichen Industriestaaten und spielen folglich in der Gesundheitsversorgung eine entscheidende Rolle. In der Bundesrepublik Deutschland liegt die Ein-Jahres-Prävalenz nach dem Bundes-Gesundheitssurvey aus dem Jahr 1998 bei 32% in der Altersgruppe von 18 bis 65 Jahren (Wittchen & Jacobi, 2001). Entsprechend erfüllt annähernd ein Drittel der erwachsenen Allgemeinbevölkerung im Laufe eines Jahres die Kriterien für das Vorliegen der Diagnose einer psychischen Störung. Die Relevanz der Verbreitung psychischer Erkrankungen liegt zum einen in dem individuellen Leidensdruck und den Folgen der emotionalen, kognitiven, körperlichen und sozialen Funktionseinschränkungen für den Patienten selbst. Zum anderen stellen sich auch erhebliche Konsequenzen auf gesellschaftlicher und ökonomischer Ebene dar: Psychische Erkrankungen gehen häufig mit verminderter Leistungsfähigkeit, erhöhten Arbeitsausfallraten und einer intensiven Nutzung des Gesundheitssystems einher (Jacobi, Klose & Wittchen, 2004).

Für die Behandlung psychischer Erkrankungen kann in Deutschland auf ein gut ausgebautes Versorgungssystem zugegriffen werden. Dabei zeichnet sich die Versorgungsstruktur durch ein, neben der ambulanten und teilstationären Behandlung, besonders dicht ausgebautes System der stationären Versorgung aus (Schulz, Barghaan, Harfst, Dirmaier, Watzke & Koch, 2006). Der Schwerpunkt der Behandlung liegt bei einem großen Teil der stationären Versorgung auf psychotherapeutischen Verfahren. Empirische Befunde zur Wirksamkeit der stationären psychotherapeutischen Behandlung deuten generell, wenn auch durch ein unterschiedlich hohes Ausmaß an empirischen Befunden belegbar, auf eine gute Wirksamkeit hin. Dennoch zeigt sich eine gewisse Varianz, anscheinend profitieren einige Patienten(-gruppen) besser von den Behandlungen als andere. Einen Schwerpunkt in der Wirksamkeitsforschung der psychotherapeutischen Versorgung stellt die Frage nach Merkmalen und Faktoren, die das Behandlungsergebnis vorhersagen dar (Clarkin & Levy, 2004). Nur wenn anhand bestimmter personenbezogener Merkmale festgestellt werden kann, welche Patienten weniger von der Behandlung profitieren, kann eine individualisierte Anpassung und Optimierung des Behandlungsprozesses stattfinden. Ebenso kann die Effektivität der Behandlung verbessert werden, wenn fundierte Erkenntnisse dazu vorliegen, wie bestimmte Prozesse im Behandlungsverlauf mit dem Behandlungsergebnis zusammenhängen. Prozessmerkmale beziehen sich dabei auf die Dauer, den Umfang und die Art der erhaltenen Behandlung. Letztlich können auch Veränderungen der Strukturen (z.B. personelle und räumliche Ausstattung) besonders dann zu effektiveren Behandlungen

führen, wenn eine Vorhersage des Ergebnisses durch Strukturmerkmale empirisch begründet ist.

Zur Vorhersage des Behandlungsergebnisses anhand von Patienten-, Prozess- und Strukturmerkmale lässt sich bislang kein einheitliches Bild vorfinden (Beutler, Malik, Alimohamed, Harwood, Talebi, Noble et al., 2004; Clarkin & Levy, 2004; Lambert & Ogles, 2004; Orlinsky, Ronnestad & Willutzki, 2004). Während Patientenmerkmale häufiger untersucht wurden, allerdings mit zum Teil inkonsistenten Ergebnissen (Clarkin & Levy, 2004), ist die Forschungsaktivität zur Vorhersage des Behandlungserfolgs anhand von Prozessmerkmalen geringer ausgeprägt (Orlinsky et al., 2004). Aus den Studien, in denen der Zusammenhang von Behandlungsdauer und -wirksamkeit untersucht wird, lassen sich keine einheitlichen Schlussfolgerungen ableiten (Lambert & Ogles, 2004; Steffanowski, Löschmann, Schmidt, Wittmann & Nübling, 2007). Neben der Dauer steht in jüngster Zeit auch die Untersuchung von Veränderungen im Verlauf der Behandlung im Fokus der Forschung (Lambert, 2005; Lutz, Stulz & Köck, 2009; Tadic, Helmreich, Mergl, Hautzinger, Kohnen, Henkel et al., 2010). Erkenntnisse über Veränderungen im Behandlungsverlauf können zu einer frühzeitigen Anpassung und Optimierung der Behandlung beitragen.

Aus der Forschungsperspektive stellt die Untersuchung der Veränderungen im Laufe der Behandlung eine hohe Relevanz im Sinne eines Erkenntnisgewinns über Behandlungsprozesse dar: Bislang wurde in fast allen Studien zur Wirksamkeit der stationären Behandlung ausschließlich die Situation zu Beginn und bei Beendigung der Behandlung, zum Teil auch katamnestisch, erhoben (Härter, Sitta, Keller, Metzger, Wiegand, Schell et al., 2004; Kösters, Nachtigall, Burlingame & Strauß, 2006; Steffanowski et al., 2007). Welche Veränderungen, Schwankungen oder Stagnationen *während* der Behandlung im patientenorientierten Outcome auftreten, ist weitgehend unbekannt. Ein Forschungsfokus richtet sich dabei auf die Identifikation von Patientensubgruppen, die distinkte Verlaufsmuster während der Behandlung aufweisen. Die empirische Untersuchung von Symptomverläufen wird verstärkt erst in den letzten Jahren von wenigen Forschungsgruppen verfolgt, unter anderem auch bedingt durch die Möglichkeiten neuerer statistischer Verfahren zur Verlaufsanalyse (Keller & Hautzinger, 2007; Lutz, Stulz, Smart & Lambert, 2007b; Stulz, Lutz, Leach, Lucock & Barkham, 2007; Stulz, Thase, Klein, Manber & Crits-Christoph, 2010).

Ziel der vorliegenden Arbeit ist die Untersuchung der Vorhersagekraft des Verlaufs der Symptombelastung während der stationären Behandlung für das längerfristige Behandlungsergebnis. Der Symptomverlauf wird über wöchentliche Erhebungen im Verlauf der Behandlung abgebildet. In der *ersten Fragestellung* wird untersucht, ob sich Patientengruppen finden lassen, die innerhalb einer Gruppe möglichst homogene und zwischen den Gruppen möglichst heterogene Verläufe der Symptombelastung aufweisen.

Dafür erfolgt in einem ersten Schritt eine Verlaufsanalyse zur Identifikation von distinkten Symptomverlaufgruppen mittels einer explorativen methodischen Vorgehensweise. Zusätzlich wird anhand der gleichen Stichprobe eine aus theoretisch-konzeptionell abgeleiteten Kriterien a priori definierte Einteilung von Verlaufgruppen vorgenommen. Es wird untersucht, ob und welche Verläufe der Symptombelastung mit der explorativen und der theoriegeleiteten Vorgehensweise aufzufinden sind.

In der *zweiten Fragestellung* wird untersucht, welche Patientenmerkmale eine Vorhersage des Symptomverlaufs während der Behandlung erlauben. Aufgrund der geringen Anzahl von Studien zur Identifikation und Vorhersage von Symptomverläufen stützt sich die Auswahl der untersuchten Merkmale auf bisherige empirische Befunde sowohl zur Vorhersage des Behandlungsergebnisses, als auch zur Vorhersage der Behandlungsdauer. Der Verlauf der Symptombelastung kann auch, im Sinne einer hochfrequenten Erfassung des Ergebnisses über die gesamte Dauer der Behandlung, als eine Integration von Dauer und Ergebnis verstanden werden (Kendall, Holmbeck & Verduin, 2004).

Die *dritte Fragestellung* stellt die Hauptfragestellung der Arbeit dar. Es soll untersucht werden, inwiefern die identifizierten Verlaufgruppen eine Vorhersage des längerfristigen Behandlungsergebnisses ermöglichen. Das Behandlungsergebnis wird über das primäre Outcome *Symptombelastung* und das sekundäre Outcome *psychische Aspekte der gesundheitsbezogenen Lebensqualität* abgebildet. Patientenmerkmale und weitere Prozessmerkmale, die potentiell einen Zusammenhang mit dem Behandlungsergebnis aufweisen, werden dabei im Sinne der Kontrolle von Störvariablen berücksichtigt. Hiermit sollen potentielle Zusammenhänge oder Wechselwirkungen zwischen Symptomverlauf und empirisch fundierten Patienten- sowie Prozessmerkmalen Berücksichtigung finden. Alle Analysen werden sowohl anhand der explorativ gefundenen Verlaufgruppen, als auch an den theoretisch-konzeptuell festgelegten Verlaufgruppen durchgeführt. Abschließend werden die beiden Verlaufgruppeneinteilungen hinsichtlich ihres Vorhersagewertes für das längerfristige Behandlungsergebnis miteinander verglichen.

Die Arbeit ist in fünf Kapitel untergliedert. Im ersten Kapitel werden die theoretisch-empirischen Grundlagen der Untersuchungsfragestellungen dargestellt. Um die Relevanz der Fragestellungen zu verdeutlichen und zur besseren Einordnung der Ergebnisse werden zuerst die Versorgungssituation psychisch erkrankter Menschen sowie die methodischen Aspekte der Versorgungs- und Psychotherapieforschung erläutert. Anschließend werden die empirischen Befunde zur generellen Wirksamkeit sowie zur Vorhersage der Wirksamkeit stationärer Psychotherapie betrachtet. Ein Schwerpunkt liegt dabei auf der Darstellung der empirischen Grundlage zur Identifikation von Verläufen sowie dem Zusammenhang von Behandlungsverlauf und Behandlungsdauer mit dem Behandlungsergebnis. Zur Identifikation potentieller Prädiktoren des Behandlungsverlaufs wird abschließend ein Überblick über den

Forschungsstand zur Prädiktion der Behandlungsdauer gegeben. Im zweiten Kapitel werden die Fragestellungen und Untersuchungshypothesen formuliert. Das dritte Kapitel beinhaltet die Methodik der Untersuchung: Neben der Darstellung des Studiendesigns, der Untersuchungsdurchführung und der Operationalisierung der untersuchten Merkmale erfolgt eine detaillierte Beschreibung der angewendeten statistischen Verfahren. Das vierte Kapitel enthält die Darstellung der Ergebnisse. Im fünften und letzten Kapitel werden die Ergebnisse im Kontext des aktuellen Forschungsstands diskutiert. Neben der inhaltlichen Diskussion erfolgt eine Auseinandersetzung mit methodischen Aspekten der Untersuchung. Abschließend wird ein Ausblick über forschungsrelevante und klinische Implikationen dargestellt, die aus den Ergebnissen abgeleitet werden können.

2 Theoretischer Hintergrund

Im ersten Kapitel der vorliegenden Arbeit werden die theoretischen Grundlagen zu den im zweiten Kapitel formulierten Fragestellungen dargestellt. Zur Einordnung der Fragestellungen und Ergebnisse wird zuerst ein kurzer Abriss zur Versorgungsstruktur für Menschen mit psychischen Erkrankungen dargestellt (Abschnitt 2.1). Dabei liegt der Schwerpunkt auf dem untersuchten Setting der stationären psychotherapeutischen Versorgung. Zur besseren Einordnung und Interpretation der empirischen Befunde aus der Psychotherapie- und Versorgungsforschung, werden in Abschnitt 2.2 methodische Besonderheiten der Versorgungsforschung und der Psychotherapieforschung dargestellt. Anschließend werden die empirischen Befunde zur Wirksamkeit von Psychotherapie umrissen (Abschnitt 2.3). Hauptaugenmerk der Untersuchung ist die Vorhersage des Behandlungsergebnisses durch Symptomverläufe und die Behandlungsdauer, bei der empirisch fundierte Prädiktoren des Outcomes berücksichtigt werden. Dementsprechend liefert der Hauptteil zur Empirie einen Überblick über empirische Befunde zur Prädiktion des Behandlungserfolgs: Neben dem Vorhersagewert von Patientenmerkmalen (Abschnitt 2.4) werden im Besonderen die Behandlungsdauer und die Symptomveränderung während der Behandlung in ihrem Vorhersagewert für das Behandlungsergebnis beleuchtet (Abschnitt 2.5). Der letzte Abschnitt des Kapitels beinhaltet den aktuellen Forschungsstand zu den patientenseitigen Prädiktoren der Behandlungsdauer und des Symptomverlaufs, die zur Formulierung von Annahmen hinsichtlich der Vorhersage von Symptomverläufen beitragen sollen (Abschnitt 2.6).

2.1 Psychotherapeutische Versorgungsstruktur in Deutschland

2.1.1 Behandlungsstrukturen

Die Bundesrepublik Deutschland verfügt über ein im internationalen Vergleich besonders dicht ausgebautes System der psychotherapeutischen Versorgung. Dies betrifft neben dem Bereich der ambulanten Versorgung, der hauptsächlich über niedergelassene ärztliche oder psychologische Psychotherapeuten abgedeckt wird, und der teilstationären Behandlung vor allem den stationären Versorgungsbereich (Schulz, Barghaan, Harfst & Koch, 2008). Dieser untergliedert sich in die folgenden drei Teilsysteme: die *psychiatrisch/psychotherapeutische Krankenhausversorgung*, die *psychosomatisch/psychotherapeutische Krankenhausversorgung (Akutpsychosomatik)* und die *medizinische Rehabilitation des Indikationsbereiches psychische/psychosomatische Erkrankungen* (Schulz et al., 2006). Eine Besonderheit besteht darin, dass sich ein substantieller Anteil der stationären Versorgung – anders als

international üblich – außerhalb der psychiatrischen Versorgung im Bereich der medizinische Rehabilitation entwickelt hat, der hauptsächlich durch die gesetzliche Rentenversicherung finanziert wird. Historisch ist der intensive Wachstumsprozess der Rehabilitation psychischer/psychosomatischer Erkrankungen in den letzten Jahrzehnten unter anderem in der verstärkten Spezialisierung auf psychotherapeutische Behandlungsmaßnahmen (in Abgrenzung zu der stark biologischen Ausrichtung der Psychiatrie der Nachkriegszeit) begründet. Zudem fand eine Erweiterung zu einem multimodalen Behandlungskonzept, in dem die im Fokus stehenden psychotherapeutischen Behandlungen durch vielfältige Bewegungs- und Kreativtherapien sowie sozialtherapeutische Versorgung ergänzt werden, statt (Koch & Potreck-Rose, 1994; Schulz et al., 2008).

Die Behandlung in der Rehabilitation psychischer/psychosomatischer Erkrankungen erfolgt nach einem auf dem biopsychosozialen Krankheitsmodell basierenden, multimodalen Ansatz, wobei die zentralen Interventionen innerhalb des ganzheitlichen Rehabilitationskonzepts die psychotherapeutischen Maßnahmen sind (Koch & Schulz, 2008). Die am häufigsten in der Rehabilitation psychischer/psychosomatischer Erkrankungen diagnostizierten psychischen Störungsbilder (ausschließlich der Suchterkrankungen) stellen mit 27% die affektiven Störungen dar. Darunter finden sich mit über 86% hauptsächlich depressive Störungen. Weitere häufige Diagnosen sind Belastungs- und Anpassungsstörungen (16%), Angst- und Zwangsstörungen (7.5%), somatoforme Störungen (7%), Persönlichkeits- und Verhaltensstörungen (3.1%) sowie Essstörungen (1.5%) (Schulz et al., 2008). Vergleiche mit den Patienten der psychotherapeutischen wie auch der psychiatrischen Krankenhäuser weisen auf sich stark überschneidende Diagnosespektren hin. Eine Spezifität der psychiatrischen Behandlung liegt in der Behandlung von Patienten mit schizophrenen Störungen (F2-Diagnosen klassifiziert nach der Internationalen Klassifikation psychischer Störungen, ICD-10; Dilling, Mombour & Schmidt, 2000), die in der Rehabilitationsbehandlung selten vorzufinden sind (Barghaan, Harfst, Watzke, Dirmaier, Koch & Schulz, 2007). Sowohl hinsichtlich der differenziellen Indikation als auch in Bezug auf die Zielsetzung ist eine klare Abgrenzung der drei Versorgungssysteme aufgrund der hohen Überlappungen schwer vorzunehmen (Schulz & Koch, 2002). Die strukturellen Besonderheiten des stationären Settings werden aktuell zum Teil sehr kontrovers diskutiert. Hier stehen unter anderem Fragen der Zuständigkeit und Eignung der Versorgungssysteme, des Aufwandes, des Behandlungserfolgs und der Kosten im Vordergrund. Ein besonders schwieriges Problem stellt dabei die Abgrenzung von Krankenhaus- und Rehabilitationsbehandlung dar, da die in der somatischen Medizin hier üblicherweise zur Abgrenzung herangezogenen Kriterien nur begrenzt auf den Bereich der psychischen Störungen übertragbar sind (für einen Überblick vgl. Koch & Schulz, 2003; Schulz & Koch, 2002).

In Abgrenzung zur ambulanten und teilstationären Psychotherapie lassen sich für alle drei stationären Settings folgende spezifischen Komponenten benennen: Das stationäre Setting bietet einen Schutzraum mit einer klar vorgegebenen Tagesstruktur, eine intensivere Dichte an Behandlungen und damit auch eine kontinuierliche Ansprechbarkeit des professionellen Helfersonals beispielsweise in Krisensituationen, und letztlich auch eine Distanz zum familiären und alltäglichen Umfeld (Zeeck, Scheidt, Hartmann & Wirsching, 2003).

2.1.2 Behandlungsprozesse

Im Folgenden werden die Behandlungsprozesse der stationären Behandlung, getrennt nach Rehabilitations- und Krankenhausbehandlung und mit einem Fokus auf der Behandlungsdauer, dargestellt.

Rehabilitationsbehandlung

Ein wesentlicher Bestandteil des multimodalen Behandlungskonzeptes der Rehabilitation psychischer/psychosomatischer Erkrankungen sind die psychotherapeutischen Behandlungen. Diese werden überwiegend in Gruppenangeboten durchgeführt und sind durch die beiden therapeutischen Richtungen verhaltenstherapeutischer und psychoanalytisch begründeter Verfahren, sowie zu einem geringeren Anteil durch die Kombination der beiden, vertreten. In den im Jahr 2006 von der Deutschen Rentenversicherung Bund herausgegebenen „*Leitlinien für die sozialmedizinische Beurteilung von Menschen mit psychischen Störungen*“ wird berichtet, dass die Mehrheit der Einrichtungen auf psychodynamischer Grundlage arbeitet, wobei das Behandlungsspektrum in jüngerer Zeit vermehrt auch um verhaltenstherapeutische Elemente erweitert wird und ein Anstieg an Einrichtungen mit kognitiv-verhaltenstherapeutischem Behandlungsschwerpunkt zu verzeichnen ist (Deutsche Rentenversicherung Bund, 2006). Dies konnte auch in umfangreichen Prozessanalysen bestätigt werden (Barghaan et al., 2007; Barghaan, Watzke, Koch & Schulz, 2005): Von über 45 000 Patienten aus Rehabilitationseinrichtungen des Indikationsbereichs psychische/psychosomatische Erkrankungen erhielten insgesamt 50% der Patienten überwiegend oder ausschließlich Leistungen der psychodynamischen Therapien², während 36% der Patienten überwiegend oder ausschließlich verhaltenstherapeutische Leistungen³ erhielten. Eine Kombinationstherapie, definiert darüber, dass der

² Unter dem Begriff *psychodynamische Verfahren* wurden analytische sowie tiefenpsychologisch fundierte Psychotherapien zusammengefasst.

³ Die Differenzierung zwischen *überwiegend oder ausschließlich verhaltenstherapeutischen (VT)* und *überwiegend oder ausschließlich psychodynamischen (PD)* Behandlungsverfahren wurde für jeden Patienten anhand des prozentualen Anteils der jeweiligen Maßnahmen, gemessen an der psychotherapeutischen Gesamtbehandlungszeit, vorgenommen. Waren beide Verfahren mit mehr als 25% anteiliger Behandlungszeit vertreten, wurde die Kategorie *Kombinationstherapie (VD/PT)* gewählt

Anteil beider Verfahren bei einem Patienten mehr als 25% betrug, bekamen 14% der untersuchten Patienten. Darüber hinaus zeichnet sich das therapeutische Angebot in der Rehabilitation durch ein breites Spektrum an zusätzlichen psychotherapienahen Behandlungsmaßnahmen, wie zum Beispiel Entspannungsverfahren, sowie weiteren psychotherapiefernen Maßnahmen, wie Sport- und Bewegungstherapien, Ergotherapie etc. aus. Art und Umfang der erhaltenen Behandlungsmaßnahmen werden in der Routedokumentation mithilfe der Klassifikation Therapeutischer Leistungen (KTL, Deutsche Rentenversicherung, 2007) dokumentiert. In den Analysen von Barghaan et al. (2005) konnte eine erhebliche Varianz zwischen den 67 einbezogenen Kliniken in der Dosis und der Frequenz sowohl psychotherapienaher als auch psychotherapieferner Maßnahmen⁴ gefunden werden.

Psychiatrische und psychosomatische Krankenhausbehandlung

In Kliniken für Psychiatrie und Psychotherapie stehen psychopharmakologische Behandlungen im Vordergrund, psycho- und sozialtherapeutische Maßnahmen werden ergänzend angewendet (Schulz et al., 2006; Watzke, Barghaan, Harfst, Koch & Schulz, 2009). Zu differenzierten Behandlungsprozessen in psychiatrischen Kliniken liegen jedoch erst wenige Studien vor. In einer der bislang umfangreichsten multizentrischen psychiatrischen Prozess-Outcome-Studien von Härter et al. (2004) konnte für die psychiatrische Depressionsbehandlung diese Tendenz bestätigt werden: Während fast alle Patienten eine medikamentöse Therapie erhielten (94%), traten zwischen den Kliniken große Unterschiede in der Häufigkeit von psychotherapeutischen Behandlungen auf. Insgesamt wurden ca. zwei Drittel aller Patienten zusätzlich psychotherapeutisch behandelt. In der Krankenhausbehandlung stellt sich die Bestimmung der Dosis psychotherapeutischer Behandlungsmaßnahmen als schwierig dar, da bislang kaum Daten aus Leistungsdokumentationen zur Verfügung stehen (Barghaan et al., 2007).

Behandlungsdauer

Die Behandlungsdauer stellt ein bedeutsames Prozessmerkmal dar, dem in der vorliegenden Untersuchung eine besondere Rolle zukommt. Die Optimierung von Behandlungsverläufen und -zeiten besitzt eine hohe Relevanz im Sinne des Anrechts des Patienten auf eine effektive und effiziente Behandlung mit schnellst- und größtmöglicher Minimierung des Leidens. In Hinblick auf die Ressourcenknappheit im deutschen Gesundheitswesen stellt gerade der Behandlungsaufwand unter gesellschaftlicher und gesundheitsökonomischer

(Barghaan et al., 2005).

⁴ Eine detaillierte Beschreibung der Einordnung der über die KTL-kodierten Maßnahmen in *psychotherapienah* und *psychotherapiefern* findet sich im Abschnitt 4.3.3.

Perspektive – im Sinne einer gerechten Verteilung der vorhandenen Ressourcen – einen entscheidenden Parameter dar. Zudem spielt die Identifizierung von optimierten Behandlungszeiten auch vor dem Hintergrund der Entwicklung und Implementierung pauschalierender Vergütungssysteme (Andreas, Dirmaier, Harfst, Kawski, Koch & Schulz, 2009; Andreas, Dirmaier, Koch & Schulz, 2003; Heuft, Hildenbrand & Cuntz, 2010), und der Entwicklung von Behandlungsleitlinien (DGPPN, BÄK, KBV, AWMF, AkdÄ, BPTK et al., 2009; Härter, Klesse, Berger, Bermejo, Bschor, Gensichen et al., 2010; Koch & Schulz, 2003; Watzke, Büscher, Koch & Schulz, 2005), eine grundlegende Rolle.

Die durchschnittliche Behandlungsdauer der Rehabilitationsbehandlung psychischer/ psychosomatischer Erkrankungen lag im Jahr 2004 bei 40 Tagen (Verband Deutscher Rentenversicherungsträger, 2004). In einem von der Deutschen Rentenversicherung Bund herausgegebenen „Anforderungsprofil an stationäre Einrichtungen zur medizinischen Rehabilitation von psychosomatischen und psychischen Störungen“ wird diesbezüglich angegeben, dass sich Dauer sowie Art und Umfang der zu erbringenden Leistungen am „...individuellen Rehabilitationsbedarf der Rehabilitanden/Rehabilitandinnen orientiert“. Ferner wird ausgeführt, dass „...zur Erreichung des Rehabilitationsziels in den überwiegenden Fällen ein Zeitraum von 5 bis 6 Wochen veranschlagt wird“ und „...bei Vorliegen bestimmter individueller und sozialmedizinischer Konstellationen auch die Durchführung einer auf 4 Wochen begrenzter Rehabilitation angezeigt sein kann“. Eine Verlängerung oder Verkürzung der Rehabilitationsdauer kann dabei „...bei vorliegender medizinischer Notwendigkeit und in Abhängigkeit von der Prognose im Rahmen der den Einrichtungen vorgegebenen Richtwerte (Zeitbudget) erfolgen“ (Deutsche Rentenversicherung Bund, 2010, S. 5). Im stationären Setting der Psychosomatischen Medizin und Psychotherapie betrug die durchschnittliche Behandlungsdauer im selben Jahr 41.4 Tage, in Fachabteilungen für Psychiatrie und Psychotherapie wurden die Patienten durchschnittlich 24.7 Tage behandelt (Statistisches Bundesamt, 2006). Für alle drei stationären Versorgungssysteme zeigt sich über die letzten Jahre eine Verkürzung der Behandlungsdauer (Kösters et al., 2006; Richter, 2001; Steffanowski et al., 2007), die aktuell aufgrund der unzureichenden Evidenzbasierung und der unabsehbaren Konsequenzen der Verkürzung kritisch diskutiert wird (Martinson-Schittkowski & Tolzin, 2008; Spießl & Klein, 2008). Anhand der Daten der Prozess-Outcome-Analyse von Barghaan et al. (2005) konnte in eigenen unveröffentlichten Analysen ermittelt werden, dass in den 65 untersuchten Fachkliniken für die Behandlungsdauer ein, verglichen mit der Behandlungsdosis, homogeneres Bild vorzufinden ist: Über die Hälfte der Patienten erhielt eine reguläre Behandlungszeit von 4 bis 6 Wochen, und weniger als 20% überschritten die sechswöchige Dauer. Insgesamt 96% der Patienten wurden in einem Zeitraum zwischen 3 und 9 Wochen behandelt (eigene Analysen; Datensatz vgl. Barghaan et al., 2005). Die historische

Entwicklung der Behandlungsdauer weist auf einen Verkürzung und Vereinheitlichung der Behandlungen in den letzten Jahren hin. So fanden Steffanowski et al. (2007) in der bislang einzigen Metaanalyse zur Wirksamkeit der psychosomatischen Rehabilitation⁵ in dem Publikationsjahr der Primärstudie einen der stärksten Prädiktoren der untersuchten Behandlungsdauer: In älteren Studien sind längere Behandlungszeiten vorzufinden als in Studien, die in jüngerer Zeit publiziert wurden. Diese Beobachtung lässt sich auch anhand von Ergebnissen aus weiteren Primärstudien bestätigen (Neeb, Winkler, Schröder & Mestel, 2001; Zielke, 2006, 2008). Es erscheint fraglich, ob die Behandlungszeiten derzeit auf ausreichend evidenzbasierter Grundlage festgelegt werden (Nosper, 2008; Spießl & Klein, 2008). Befunde zum Zusammenhang von Behandlungsdauer und -ergebnis sowie mögliche Effekte der Verkürzung der Behandlungszeiten werden in Abschnitt 2.5 dargestellt.

2.2 Psychotherapieforschung und Versorgungsforschung

Die wissenschaftliche Beforschung des Versorgungssystems mit der Zielsetzung, eine empirische Grundlage für die Optimierung von *Zugang*, *Prozess* und *Outcome* und somit für eine bedarfs- und verteilungsgerechte Versorgung zu schaffen, wird mit dem Ansatz der *Versorgungsforschung (health care research)* verfolgt (Raspe, Pfaff, Härter, Hart, Koch-Gromus, Schwartz et al., 2010). In der vorliegenden Arbeit steht die Untersuchung der Versorgung von Patienten mit psychischen Erkrankungen im Vordergrund. Vor allem in Hinblick auf die Beforschung des Outcomes, also der Wirksamkeit von (hier psychotherapeutischen) Behandlungen, stellen sich einige inhaltliche und methodische Besonderheiten dar, die hier zur besseren Einordnung der zu berichtenden Wirksamkeitsbefunde beschrieben werden.

In der *Psychotherapieforschung* wird hinsichtlich der Wirksamkeitsforschung zwischen den methodisch unterschiedlichen Zugängen der Efficacy- und der Effectiveness-Studien unterschieden (für einen Überblick vgl. Lambert & Ogles, 2004; Lutz, 2003; Strauss & Wittmann, 2005). Bei der *Efficacy* handelt es sich um die Wirksamkeit eines bestimmten Verfahrens, die anhand eines experimentellen Designs mit randomisierter Zuweisung zu den Behandlungs- und Kontrollbedingungen untersucht wird. Kontrollbedingungen stellen aus ethischen Gründen selten streng unbehandelte Gruppen dar, vielmehr werden entweder zwei verschiedene Verfahren miteinander verglichen oder es wird ein meist innovativeres

⁵ Der historisch begründete und fachlich umstrittene Begriff *Psychosomatische Rehabilitation* wird in der vorliegenden Arbeit, aufgrund des nicht zutreffenden Eindrucks, dass in diesem Setting hauptsächlich Patienten mit psychosomatischen Störungen behandelt werden, nur dann verwendet, wenn die jeweiligen Autoren einer zitierten Publikation diesen so benennen. Andernfalls wird der Begriff *Rehabilitation psychischer/psychosomatischer Erkrankungen*, der fachlich angemessener erscheint, bevorzugt.

psychotherapeutisches Verfahren im Vergleich zur regulären Behandlung (*treatment as usual*, TAU) bzw. einer Behandlung mit unspezifischer Wirkung im Sinne eines Kontrollgruppendesigns untersucht. Ein weiterer Zugang zu Kontrollbedingungen stellt die Untersuchung von Patienten, die sich auf Wartelisten für psychotherapeutische Behandlungen befinden, dar. Ziel der Efficacy-Studien ist es, eine möglichst hohe interne Validität zu erreichen. Dabei wird versucht, mögliche Störvariablen, die mit dem Therapieergebnis zusammenhängen könnten, maximal zu kontrollieren. Die gefundenen Effekte sollen (möglichst) eindeutig auf das untersuchte Behandlungsverfahren zurückgeführt werden können (Buchkremer & Klingberg, 2001). Dahingegen beschreibt die *Effectiveness* die Wirksamkeit psychotherapeutischer Behandlungen unter realen Bedingungen. In diesem Feld der Psychotherapieforschung ist die Versorgungsforschung psychotherapeutischer Behandlung angesiedelt. Ein wesentliches Merkmal der Versorgungsforschung besteht darin, dass über vielfältige methodische Zugänge Erkenntnisse anhand realer Bedingungen und nicht über streng experimentelle Designs gewonnen werden. Durch die Erhebung unter klinisch repräsentativen Bedingungen wird die externe Validität erhöht und der Transfer klinischer Implementierungen in die Praxis erleichtert (Watzke et al., 2009). Studien aus der Versorgungsforschung werden meist als Beobachtungsstudien oder mit einem quasi-experimentellen Design durchgeführt. Aufgrund der realen Bedingungen, unter denen untersucht wird, kann in Beobachtungsstudien eine höhere externe Validität, also die Generalisierbarkeit der Ergebnisse, erreicht werden. Dennoch sollte schon bei der Auswahl der untersuchten Bedingungen (Kliniken, Settings) und der rekrutierten Patienten auf die Repräsentativität geachtet werden. Zusätzlich sollte die Übertragbarkeit der Ergebnisse durch Repräsentativitätsanalysen hinsichtlich der Patienten, die nicht an der Studie teilnahmen, abgesichert werden. Schwieriger scheint das Erreichen einer hohen internen Validität, der eindeutigen Zuordnung der gefundenen Zusammenhänge auf den untersuchten Gegenstand. Aufgrund der Untersuchung unter Realbedingungen können viele Störfaktoren auftreten, die in einem Zusammenhang mit dem untersuchten Gegenstand stehen und somit die Ergebnisse konfundieren. Diese sogenannten *Confounder* können in reinen Beobachtungsstudien nicht, wie beispielsweise in randomisiert-kontrollierten Studien, über das Untersuchungsdesign kontrolliert werden. Die Möglichkeit der Kontrolle von Confoundern besteht hier in der statistischen Berücksichtigung bzw. Adjustierung theoretische fundierter Merkmale. Durch quasi-experimentelle Designs und durch die statistische Berücksichtigung von Confounder-Variablen können kausale Interpretationen der Ergebnisse ermöglicht werden (Watzke et al., 2009). Auch hier sollte mithilfe der Untersuchung von Selektionseffekten über verschiedene Untersuchungsstichproben, der Auswahl angemessener statistischer Verfahren und ausreichend großen Untersuchungsstichproben eine möglichst hohe interne Validität

angestrebt werden (für einen Überblick über Kriterien der methodischen Studienqualität vgl. bspw. Downs & Black, 1998; Elm, Altman, Egger, Pocock, Gotsche, Vandenbroucke et al., 2007). Die Optimierung der Methoden der Versorgungsforschung wird in jüngster Zeit, z.B. durch vorgeschlagene einheitliche Standards, vermehrt angestrebt (Pfaff, Glaeske, Neugebauer & Schrappe, 2009; Raspe et al., 2010).

Die rein experimentelle Psychotherapieforschung wird wegen der in Frage gestellten Übertragbarkeit der Ergebnisse auf das naturalistische Setting seit längerer Zeit zunehmend kritisch diskutiert. Diese Entwicklung hatte zur Folge, dass in den letzten Jahrzehnten vermehrt Effectiveness-Studien durchgeführt wurden. Eine erste umfassende Meta-Analyse über 90 Studien zur Wirksamkeit von Psychotherapie, bei der gezielt die klinische Repräsentativität der Studienbedingungen einbezogen wurde, fanden die Autoren ähnliche Effektstärken für Studien unter klinisch repräsentativen Bedingungen wie für nicht-repräsentative Bedingungen (Shadish, Matt, Navarro & Phillips, 2000). In einer Studie zur Frage nach der Repräsentativität von Efficacy-Studien wurde untersucht, ob Patientenstichproben aus RCTs repräsentativ für die „reale“ Versorgungssituation sind (Stirman, Derubeis, Crits-Christoph & Rothman, 2005). Die Ergebnisse zeigen zwar hinsichtlich einiger Patientenmerkmale eine gewisse Übertragbarkeit, verdeutlichen jedoch auch, dass durch die Selektion über Ein- und Ausschlusskriterien in RCT-Studien bestimmte Patientengruppen unberücksichtigt bleiben. Diese betreffen z.B. Patienten, die sich in medikamentöser Behandlung befinden, Patienten mit komorbiden Störungen und Patienten, bei denen die zu untersuchende Diagnose vorhanden ist, jedoch nicht die Primärdiagnose ist. In einer Literaturübersicht zum Vergleich der Ergebnisse der Effectiveness und Efficacy von Psychotherapie zeigen Hunsley und Lee (2007), dass die aus Übersichtsarbeiten zur Efficacy erzielten Effektstärken auch in den Effectiveness-Studien gefunden werden. Die Autoren interpretieren ihre Ergebnisse mit einer guten Übertragbarkeit von Ergebnissen aus RCTs auf die klinische Praxis. Einschränkend muss erwähnt werden, dass sich die Autoren hauptsächlich auf den ambulanten Versorgungsbereich beziehen.

In der Psychotherapieforschung wie in der Versorgungsforschung zu psychischen Erkrankungen wird zudem die Messung nachhaltiger Effekte empfohlen (Kendall et al., 2004). Katamnestische Erhebungen dienen der Untersuchung der Stabilität und langfristigen Wirksamkeit von psychotherapeutischen Behandlungen sowie der Integration der Therapieeffekte in den Alltag. Kritisch zu betrachten sind die schwer durchführbare Kontrolle von Effekten intermittierender Ereignisse, die nach Ende der Behandlung auftreten und das langfristige Behandlungsergebnis beeinflussen können, sowie das Problem von Selektionseffekten, da in der Regel nur ein Teil der Patienten katamnestisch nachuntersucht wird (Strauss & Wittmann, 2005). Um Verzerrungen der Ergebnisse entgegenzuwirken sollten die Selektivität der Studienteilnehmer zur Katamnese überprüft werden.

2.3 Wirksamkeit stationärer psychotherapeutischer Behandlungen

Die generelle Wirksamkeit von Psychotherapie konnte in Meta-Analysen vielfach bestätigt werden (für einen Überblick vgl. Lambert & Ogles, 2004). Die Wirksamkeitsforschung von Psychotherapie bezieht sich im internationalen Rahmen allerdings hauptsächlich auf den ambulanten Versorgungsbereich, da die stationäre Versorgung außerhalb Deutschlands sowohl hinsichtlich bestehender Kapazitäten, als auch in Hinblick auf die Forschungsaktivität eine weit geringere Rolle spielt (Watzke et al., 2009). Aufgrund des gut ausgebauten stationären Systems in Deutschland liegen hier mehr Wirksamkeitsstudien für dieses Setting als für die ambulante Psychotherapie vor (Schulz et al., 2006). Im Folgenden wird erst eine Einschätzung der Messung und Bewertung von (stationärer) Psychotherapie vorgenommen und anschließend ein Überblick der empirischen Befunde zur Wirksamkeit stationärer Psychotherapie dargestellt.

2.3.1 Maße und Bewertung der Wirksamkeit stationärer Psychotherapie

Einen zentralen Aspekt der Wirksamkeitsforschung stellt die Frage dar, wie Behandlungserfolg gemessen wird und wie Erfolgskriterien bewertet werden. Die Definition des Erfolgskriteriums spielt eine bedeutende Rolle, da Ergebnisse zur Effektivität von Psychotherapie von dem angewendeten Maß abhängen und eine Vergleichbarkeit von Ergebnissen verschiedener Studien nur unter Berücksichtigung der Vergleichbarkeit der angesetzten Maßstäbe des Erfolgs möglich ist (Hill & Lambert, 2004). In einer umfassenden Übersichtsarbeit zeigen Hill und Lambert eindrücklich, dass in 384 Studien zur Wirksamkeit ambulanter Psychotherapie insgesamt 1 430 verschiedenen Messinstrumenten identifiziert werden konnten, wobei 840 dieser Messinstrumente jeweils nur einmal eingesetzt wurden (Hill & Lambert, 2004). Dabei erstreckt sich die Vielfalt über standardisierte und nicht-standardisierte Instrumente, die den Erfolg aus verschiedenen Perspektiven (bspw. aus der Perspektive des Patienten oder aus der Perspektive des Therapeuten) anhand generischer (bspw. Lebensqualität) oder spezifischer (bspw. störungsspezifischer) Maße messen. Inhaltlich stellen neben patientenseitigen Ergebniskriterien auch sozialgesellschaftliche sowie ökonomische Kriterien Ergebnisparameter dar. In einer Meta-Analyse zur Wirksamkeit der stationären psychosomatischen Rehabilitationsbehandlung konnten in 65 Studien insgesamt 531 unterschiedliche Ergebniskriterien identifiziert werden (Steffanowski et al., 2007). Diese beziehen sich sowohl auf subjektive Maße wie psychisches, körperliches oder allgemeines Befinden sowie kognitive, soziale und funktionale Fähigkeiten, als auch auf objektive und kostenrelevante Angaben wie z.B. Arbeitsunfähigkeitszeiten oder die Inanspruchnahme des Gesundheitssystems. Von den psychometrischen Testverfahren wurden am häufigsten Instrumente zur Messung der Symptombelastung eingesetzt: In 22

Studien kam die Symptom-Checkliste SCL-90-R (Franke, 2002) und in 12 Studien das Beck-Depressions-Inventar (BDI; Beck, Ward, Mendelson, Mock & Erbaugh, 1961; Hautzinger, Bailer, Worall & Keller, 1994) zur Anwendung.

Begründet ist die vorzufindende Vielfalt der Outcome-Parameter unter anderem in der Komplexität des zu erfassenden Konstrukts sowie in der Diversität der Einschätzung von Zielkriterien der psychotherapeutischen Behandlung. Für den Einsatz von psychometrischen Instrumenten sollten vor allem die gängigen Testgütekriterien *Objektivität*, d.h. die Unabhängigkeit von dem Versuchsleiter, *Reliabilität*, d.h. die Messgenauigkeit des Instruments und die *Validität*, d.h. die Gültigkeit mit der das gemessen wird, was gemessen werden soll, weitgehend erfüllt sein (Hill & Lambert, 2004).

Um Ergebnisse, die auf unterschiedlichen Messinstrumenten mit unterschiedlichem Skalenniveau basieren, vergleichen zu können, haben sich über die letzten Jahrzehnte verschiedene standardisierte Effektstärkemaße etabliert. Diese werden zumeist ermittelt, indem die Mittelwertsdifferenz von zwei untersuchten Gruppen oder, bei einem längsschnittlichen Ein-Gruppen-Prä-Post-Design, die Mittelwertsdifferenz zwischen der Prä- und der Post-Messung an einem Streuungsmaß relativiert werden, wobei hier unterschiedliche Maße verwendet werden können (vgl. auch Abschnitt 4.5.4). In Studien mit einem Kontrollgruppendesign wird die Mittelwertsdifferenz häufig an der Standardabweichung der Kontrollgruppe relativiert. Hinsichtlich des zu verwendenden Streuungsmaßes bei Ein-Gruppen-Prä-Post-Designs herrscht hingegen keine Einigkeit⁶. Da die Ergebnisse jedoch in Abhängigkeit von dem Wert der Streuung stark variieren, sollte das gewählte Streuungsmaß bei der Interpretation sowie beim Vergleich von Effektstärken berücksichtigt werden (Maier-Riehle & Zwingmann, 2000).

Die Erhebung von Veränderungen kann über die direkte oder die indirekte Veränderungsmessung erfolgen (Kohlmann & Raspe, 1998). Während die direkte Veränderungsmessung anhand einer retrospektiven Einschätzung der Veränderung von Therapiebeginn bis -ende erfolgt, wird bei der indirekten Veränderungsmessung die Differenz (im fokussierten Kriterium) zwischen Einschätzungen zu Beginn und zum Ende der Therapie gebildet. Beide Verfahren weisen methodische Vor- und Nachteile auf und werden kritisch diskutiert (Stieglitz & Baumann, 2001). Die direkte Veränderungsmessung stellt aufgrund der Einzeitpunktbefragung die ökonomischere Variante dar, allerdings weist diese auch wesentliche Einschränkungen auf: Untersuchungen konnten zeigen, dass die befragten Patienten ihr Urteil primär auf den Ist-Zustand zum Befragungszeitpunkt und aufgrund der kognitiven Überforderung weniger auf den Zeitraum der Veränderung beziehen (Guyatt, Norman, Juniper & Griffith, 2002; Schmitt & Di Fabio, 2005; van Stel, Maille, Colland &

⁶ Eine detaillierte Beschreibung der verschiedenen Möglichkeiten zur Berechnung von Prä-Post-Effektstärken ist in Abschnitt 4.5.4 zu finden.

Everaerd, 2003). Des Weiteren besteht die Möglichkeit, dass das Veränderungsurteil eher die Folge impliziter Theorien zur Erklärung der Veränderungen abbildet als die tatsächliche Veränderung (Farin, Dudeck, Meffert, Glattacker, Jäckel, Beckman et al., 2009). Zudem ist die direkte Veränderungsmessung bei mehr als zwei zu vergleichenden Zeitpunkten ungeeignet (Kohlmann & Raspe, 1998). Einschränkungen der indirekten Veränderungsmessung stellen zum einen die Validität der Messungen dar, da Verschiebungen im Bezugsrahmen der Befragten zwischen den Messungen möglich sind. Zum anderen ist die Reliabilität der Differenzwerte eingeschränkt, da mehrfache Messungen immer auch mit mehreren unabhängigen Fehlerquellen einhergehen. Letztlich hängt die Möglichkeit, Veränderungen zu erfassen, von der Änderungssensitivität der eingesetzten Skala ab (Nübling, Steffanowski, Wittmann & Schmidt, 2004). Treten Veränderungen über die Zeit auf, kann über die Differenzwerte die statistische Signifikanz der Veränderung geprüft werden. Hierbei wird jedoch nur die Bedeutsamkeit in Abgrenzung einer durch den Zufall aufgetretenen Veränderung beschrieben; inwiefern diese Veränderungen eine klinische Bedeutsamkeit erlangen, kann durch die statistische Testung nicht ermittelt werden (Kendall et al., 2004). Als Erweiterung der statistischen Signifikanz wurden Differenzwertberechnungen unter Berücksichtigung der klinischen Relevanz vorgeschlagen. Einer der bedeutendsten Vorschläge ist das von Jacobson und Truax (1991) entwickelte Konzept der *kritischen Differenz (Reliable Change Index, RCI)*. Der RCI gibt das Ausmaß einer statistisch bedeutsamen Veränderung unter Berücksichtigung der Streuung einer repräsentativen Stichprobe sowie der Messgenauigkeit des verwendeten Instrumentes an (vgl. auch Abschnitt 4.5.4). Eine klinisch signifikante Verbesserung ist erreicht, wenn sowohl der RCI überschritten wurde, als auch eine Verbesserung aus dem klinisch auffälligen in den klinisch unauffälligen Bereich (anhand des an Referenzstichproben ermittelten Cut-Off-Wertes) erfolgte. Entsprechend resultieren aus der Berechnung des RCI qualitative Ergebnisse, während das Ergebnis bei der Differenzwertberechnung als metrische Variable vorliegt.

2.3.2 Empirische Befunde zur Wirksamkeit stationärer Psychotherapie

Im deutschsprachigen Raum ist die größte Forschungsaktivität zur Wirksamkeit stationärer psychotherapeutischer Behandlung im Bereich der Rehabilitation psychischer/ psychosomatischer Erkrankungen zu finden. Entsprechend beziehen sich viele der im Folgenden beschriebenen Ergebnisse auf dieses Setting. Die Wirksamkeit der stationären Behandlung psychischer Erkrankungen wurde durch zahlreiche Programm-evaluationsstudien (vgl. z.B. Broda, Bürger, Dinger-Broda & Massing, 1996; Dilcher, Mestel, Klingelhöfer, Köbel, Sprenger & Stauss, 2000; Schulz, Lotz-Rambaldi, Koch, Jürgensen & Rüdell, 1999; Tritt, von Heymann, Loew, Benker, Bleichner, Buchmueller et al., 2003; Zielke,

1993) sowie durch indikationsspezifische Untersuchungen im Akut- und Rehabilitationssetting (vgl. z.B. Härter et al., 2004; Hillert, Maasche, Kretschmer, Ehrig, Schmitz & Fichter, 1999; Kächele, Kordy & Richard, 2001) überprüft. Dabei zeigen sich sowohl für verschiedene generische und spezifische Parameter (Lebensqualität, Wohlbefinden, Symptomverbesserung) als auch für kostenrelevante Merkmale (Arbeitsunfähigkeitszeiten, ambulante Arztbesuche, Krankenhausaufenthalte, Medikamentenkonsum) positive Behandlungsergebnisse (Koch & Schulz, 2008). Diese Ergebnisse lassen sich in einigen Studien zudem auch katamnestisch noch zeigen, d.h., dass auch noch ein halbes oder ein Jahr nach Verlassen der Klinik statistisch signifikante und klinisch relevante Verbesserungen in einzelnen zentralen Ergebniskriterien im Vergleich zur Baseline-Erhebung vorliegen (z.B. Barghaan et al., 2005; Schmidt, Nübling & Wittmann, 2000). In der bereits erwähnten Metaanalyse zu den Effekten⁷ stationärer psychosomatischer Rehabilitation (Steffanowski et al., 2007) sind diese Ergebnisse sowohl für Prä-Post-Vergleiche als auch katamnestisch weitgehend wieder zu finden. Für insgesamt 54 Primärstudien ließ sich eine gemittelte und gewichtete Gesamteffektstärke der Veränderung von der Aufnahme zur Entlassung über alle Studien und Ergebnismaße von $d = 0.51$ finden. Die Effektstärken der Ergebnisse der Primärstudien wurden dabei an der Prä-Streuung standardisiert. In 46 Studien wurden auch katamnestische Ergebnisse berichtet, die im Durchschnitt ein Jahr nach der Entlassung erhoben wurden. Die gemittelte gewichtete Gesamteffektstärke für die Veränderung von der Aufnahme zur Katamnese betrug 0.41. Eine indikationsspezifische Differenzierung zeigte deutliche Unterschiede in den Effektstärken, wobei Patienten mit depressiven Störungen am stärksten von der rehabilitativen Behandlung profitieren (Effektstärke zum Entlasszeitpunkt: $d = 0.84$; Effektstärke zum Katamnese-Zeitpunkt: $d = 0.67$). Betrachtet man ausschließlich die Ergebnisse reliabler und valider Depressionsinventare, z.B. anhand des Beck-Depressions-Inventars (BDI; Hautzinger et al., 1994), erhöhen sich die Effektstärken bei depressiven Patienten zur Entlassung sogar auf $d = 1.07$ und zur Katamnese auf $d = 0.76$. Dahingegen verbesserten sich Patienten mit somatoformen Störungen im geringsten Ausmaß (Gesamteffektstärke zur Entlassung: $d = 0.49$; Gesamteffektstärke zur Katamnese: $d = 0.33$). Im mittleren Bereich liegen die Effektstärken der Patienten mit Angststörungen (Gesamteffektstärke zur Entlassung: $d = 0.71$; Gesamteffektstärke zur Katamnese: $d = 0.45$). Für die beiden letztgenannten Störungsgruppen finden sich keine großen Unterschiede zwischen den generischen und den

⁷ Die Begriffe Effekt und Effektstärke werden in vielen Studien synonym verwendet und beschreiben meist die durch eine Behandlung erzielten Veränderungen, ausgedrückt durch ein Effektstärkemaß. In der vorliegenden Arbeit wird der Begriff Effektstärke bevorzugt, da hiermit eine genauere Beschreibung der gefundenen Veränderungsgröße erfolgt. Von Effekten einer Behandlung kann im engeren Sinne nur dann gesprochen werden, wenn diese tatsächlich auf die Intervention zurück geführt werden können. Dies ist v.a. in Beobachtungsstudien nur schwer sicherzustellen (vgl. Abschnitt 2.2).

störungsspezifischen Outcome-Maßen (Steffanowski et al., 2007). Effektstärken in ähnlicher Größenordnung sind auch in einigen multizentrischen Studien mit Primärdaten zu finden. Barghaan et al. (2005) untersuchten das katamnestische Behandlungsergebnis von 3 113 Patienten 6 Monate nach der stationären Rehabilitationsbehandlung. Im Bereich der psychischen Aspekte gesundheitsbezogener Lebensqualität, operationalisiert über die psychische Summenskala des Short-Form 12 Health Survey (SF-12, Kurzversion des SF-36, Ware, 2000; vgl. auch Abschnitt 4.3.2), zeigten sich Verbesserungen zur Katamnese mit einer durchschnittlichen Effektstärke von $d = 0.80$. In der Symptombelastung, gemessen mit der General Severity Index (GSI) der Symptom Checklist SCL-9 (Klaghofer & Brähler, 2001), einer Kurzform der SCL-90-R (Franke, 2002), wurde eine durchschnittliche Effektstärke von $d = 0.44$ erreicht. Die Effektstärkenberechnung basierte dabei auf der Standardisierung der Mittelwertsdifferenzen an der Streuung der Prä-Messungen.

In einer multizentrischen Studie von Tritt et al. (2003) mit einer Stichprobe von knapp 6 000 Patienten aus 11 Fachkliniken (Krankenhausbehandlung Psychosomatik und Psychotherapie) wurde die Symptombelastung ebenfalls anhand des GSI der SCL-90-R (Franke, 2002) erhoben. Die ermittelte Prä-Post-Effektstärke lag bei $d = 0.83$, katamnestische Ergebnisse lagen allerdings nicht vor. Eine Übersicht über internationale Studien sowie spezifisch über deutschsprachige Studien der Krankenhausbehandlung Psychosomatik und Psychotherapie, liefern Kösters et al. (2006) und Kösters und Strauß (2007). In der Übersichtsarbeit zu den internationalen Studien finden die Autoren über den Zeitraum von fast 25 Jahren 46 Studien im Prä-Post-Design und weitere 24 Studien, die unter kontrollierten Bedingungen durchgeführt wurden. Bei der Berechnung der Effektstärken im Kontrollgruppendesign wurde die Standardabweichung der unbehandelten Kontrollgruppe verwendet, für die Prä-Post-Effektstärken wählten die Autoren die Streuung der Prä-Werte. Die Prä-Post-Gesamteffektstärke beträgt für die Studien im Kontrollgruppendesign $d = 0.31$ und für die Studien im Eingruppen-Prä-Post-Design $d = 0.59$. Auch bei alleiniger Betrachtung der eingeschlossenen (und einiger im folgenden Jahr hinzugekommen) deutschen Studien, die ausschließlich im Prä-Post-Design vorzufinden waren, fanden die Autoren eine korrespondierende Gesamteffektstärke von $d = 0.58$ (Kösters & Strauß, 2007).

Im Rahmen eines Qualitätssicherungsprogramms untersuchten Härter et al. (2004) in 24 psychiatrisch-psychotherapeutischen Fachkliniken über 3 000 Patienten mit depressiven Störungen. Dabei wiesen die Effektstärken in zwei störungsspezifischen Instrumenten hohe Werte auf: Im BDI (Hautzinger et al., 1994) erreichten die Patienten bis zur Entlassung eine Effektstärke von $d = 1.5$, auf der Hamilton Depressionsskala (Hamilton Rating Scale for Depression HAMD; Hamilton, 1960) sogar eine mittlere Effektstärke von $d = 2.3$. Die über den HAMD erzielten hohen Effektstärken können zum einen auf die Beurteilerperspektive zurückgeführt werden. Anders als beim BDI handelt es sich bei dem HAMD um ein

Fremdbeurteilungsinstrument, sodass die Einschätzung der Symptombelastung von den behandelnden Ärzten vorgenommen wurde. Auch Hautzinger, De Jong-Meyer, Treiber, Rudolf und Thien (1996a) finden in einer indikationsspezifischen Wirksamkeitsstudie, die auch Patienten aus der stationären psychiatrisch-psychotherapeutischen Behandlung einschließt, ähnlich hohe Effektstärken im BDI ($d = 2.10$) und HAMD ($d = 2.50$). Weitere Erklärungsansätze für die Unterschiede zu den Effektstärken aus den Studien aus nicht-psychiatrischen stationären Versorgungsbereichen können in mehrfacher Hinsicht erwogen werden: Zum einen stellt die Untersuchungsstichprobe von Härter et al. (2004) aufgrund des eingeschränkten Diagnosespektrums eine homogenere Stichprobe dar. Entsprechend konnten störungsspezifische Instrumente eingesetzt werden, die zusätzlich die Varianz verringern können. Eine geringere Varianz wirkt sich direkt auf die Höhe der Effektstärke aus. Zum anderen ziehen die Autoren in Erwägung, dass die vergleichsweise hohen mittleren Behandlungsdauern (durchschnittlich 63.6 Tage) mit dem Outcome zusammenhängen könnten. Bei einem großen Teil der untersuchten Stichprobe handelte es sich um schwer oder chronisch depressiv erkrankte Patienten, die durch ihre hohe Ausgangsbelastung rein messmethodisch größere Veränderungen erzielen können als weniger schwer erkrankte Patienten. So findet auch Nosper (1999b) in Subgruppenanalysen an einer Stichprobe schwer erkrankter Patienten aus der Rehabilitation psychischer/psychosomatischer Erkrankungen in der störungsspezifischen Allgemeinen Depressionsskala (ADS; Hautzinger & Bailer, 1993) Effektstärken von $d = 1.70$, die weit höher liegen als die der Gesamtstichprobe ($d = 0.80$). Damit einhergehend kann auch die Tatsache, dass im psychiatrisch-psychotherapeutischen Setting fast alle Patienten pharmakotherapeutisch mit Antidepressiva (94%) behandelt wurden, mit den hohen Effektstärken zum Entlasszeitpunkt zusammenhängen (Härter et al., 2004). Zur Einschätzung der Stabilität der Veränderungen sollten Katamnese-Erhebungen durchgeführt werden.

Zusammenfassend zeigen sich, wie in der international weiter verbreiteten ambulanten Psychotherapieforschung, auch für die Wirksamkeit der stationären psychotherapeutischen Behandlung konsistent mittlere bis hohe Effektstärken. Generell fallen diese störungsspezifisch unterschiedlich hoch aus und vor allem Patienten mit depressiven Störungen und Patienten mit Angsterkrankungen weisen in störungsspezifischen Outcome-Maßen hohe Effektstärken auf (Härter et al., 2004; Kösters et al., 2006; Steffanowski et al., 2007). Inwiefern die erzielten Effektstärken kausal auf Behandlungseffekte zurückzuführen sind, hängt von der Güte des jeweiligen Studiendesigns und den angewendeten Methoden ab. Einschränkend muss darauf hingewiesen werden, dass lediglich in der Meta-Analyse von Steffanowski et al. (2007) und in den Primärdaten-Analysen von Barghaan et al. (Barghaan et al., 2005) längerfristige Ergebnisse berichtet werden. Beide Analysen beziehen sich auf

Studien bzw. Daten der Rehabilitationsbehandlung psychischer/psychosomatischer Erkrankungen. Für das akutstationäre Setting liegen bislang noch keine katamnesticen Effektstärken aus multizentrischen Studien oder Übersichtsarbeiten vor.

2.4 Vorhersage der Wirksamkeit stationärer Psychotherapie

Obwohl die generelle Wirksamkeit der stationären psychotherapeutischen Behandlung durch viele Untersuchungen belegt ist, besteht noch viel Unklarheit darüber, durch welche Merkmale der Behandlungserfolg vorhergesagt werden kann. Um die Patienten zu identifizieren, die voraussichtlich ein besseres oder schlechteres Behandlungsergebnis erzielen, können zum einen patientenseitige Merkmale, die schon zu bzw. vor Beginn der Behandlung feststehen, herangezogen werden. Weiterhin können auch Merkmale, die aus dem Prozess der Behandlung hervorgehen, in einem bedeutenden Zusammenhang mit dem Behandlungsergebnis stehen. Merkmale, die sich auf die strukturellen Bedingungen der Behandlung beziehen, wie etwa Merkmale der Institution, der Behandlung oder der Behandler beziehen (Personalschlüssel, Raumgröße und -ausstattung, Medien etc.), bilden die dritte Gruppe an potentiellen Prädiktoren des Behandlungsergebnisses.

In den folgenden Abschnitten werden empirische Befunde zur Vorhersage des Behandlungserfolgs dargestellt. Zuerst werden Ergebnisse der Vorhersage durch Patientenmerkmale beschrieben, anschließend folgt eine Darstellung der Befunde zur Vorhersage aus Prozessmerkmalen. Hier liegt der Fokus auf den Merkmalen Dauer bzw. Dosis und Symptomverlauf. Weitere Prozess-, Therapeuten- sowie Strukturmerkmale werden hier nicht im Einzelnen erläutert, da sie nicht im Fokus der Arbeit stehen.

2.4.1 Patientenmerkmale als Prädiktoren des Behandlungserfolgs

Patientencharakteristika, die in einem potentiellen Zusammenhang mit dem Behandlungsergebnis oder mit dem Behandlungsprozess stehen können, sind vielfältig und in ihrer Anzahl nahezu unbegrenzt (Clarkin & Levy, 2004). Diese reichen theoretisch von genetisch-biologischen Faktoren, über soziodemographische und klinische Merkmale, stabile oder variable personenbezogene Eigenschaften bis hin zu den die Person betreffende Umweltbedingungen. Im Folgenden werden die relevanten Befunde zu soziodemographischen (Alter, Geschlecht, Familienstand, sozioökonomischer Status), klinischen (Symptomschwere, Diagnose, Chronifizierung) und untersuchten Persönlichkeitsmerkmalen dargestellt. Dabei sollen neben den zusammenfassenden Befunden aus der ambulanten Psychotherapieforschung jeweils vor allem auch Ergebnisse zur Vorhersage des Ergebnisses von stationären Behandlungen betrachtet werden. Ein

zusammenfassender Überblick zur Bedeutung von patientenseitigen Prädiktoren in der ambulanten Psychotherapie ist bei Clarkin und Levy (2004) zu finden.

Soziodemographische Merkmale

Hinsichtlich des *Alters* lässt sich in Meta-Analysen kein eindeutiger Zusammenhang zum Behandlungsergebnis ambulanter Psychotherapie finden: Studien, die Stichproben älterer Depressionspatienten einbeziehen, resultieren in ähnlichen Effektstärken wie Studien mit durchschnittlichen Altersstichproben (Clarkin & Levy, 2004). Andererseits formulieren Beutler, Castonguay und Follette (2006)⁸, basierend auf einer Literaturzusammenfassung, dass das Alter bei depressiven Erkrankungen einen negativen Prädiktor des generellen Psychotherapie-Outcomes darstellt. Widersprüchliche Ergebnisse zeigen sich auch im stationären Setting. Hier finden Barghaan et al. (2005) in den Analysen des indikationsübergreifenden Datensatzes mit mehr als 3 000 Patienten bei älteren Patienten ein schlechteres Outcome im Vergleich zu jüngeren Patienten. In der Metaanalyse von Steffanowski et al. (2007) lässt sich hingegen kein Zusammenhang zwischen Alter und Behandlungsergebnis feststellen. Das Alter scheint über verschiedene Variablen moderiert zu werden. So berichten Clarkin und Levy (2004), dass ein höheres Alter bei depressiven Patienten nur im Zusammenhang mit körperlichen Funktionseinschränkungen mit einem schlechteren Behandlungsergebnis einhergeht. Auch Barghaan et al. (2005) finden altersabhängige Unterschiede in dem Prozessmerkmal Dosis psychotherapienaher Behandlungen. So können die schlechteren Ergebnisse bei älteren Patienten möglicherweise durch die geringere Dosis an Psychotherapie als singular über das Lebensalter erklärt werden. Allerdings zeigt eine kürzlich veröffentlichte multizentrische Studie zur Vorhersage von Non-Response in der stationär-psychiatrischen Depressionsbehandlung, dass sowohl univariat als auch multivariat – unter Berücksichtigung weiterer Patienten- und Prozessmerkmale – kein Zusammenhang zwischen Alter und Behandlungserfolg besteht (Hölzel, Wolff, Kriston & Härter, 2010). Einheitlicher sind die Ergebnisse zum *Geschlecht*. Sowohl in den Übersichtsarbeiten aus dem ambulanten Setting, als auch in den oben erwähnten Studien steht das Geschlecht in keinem bedeutenden Zusammenhang zum Outcome. Dennoch sollten die unterschiedlichen Prävalenzen einiger psychischer Erkrankungen (z.B. bei depressive Störungen und Essstörungen) hinsichtlich der Geschlechterverteilung bei der Untersuchung von Prädiktoren des Behandlungserfolgs

⁸ Die Autoren entwickelten im Rahmen einer „Task Force die Principles of Therapeutic Change“ eine Auflistung von Grundsätzen der effektiven psychotherapeutischen Behandlung, basierend auf empirischen Ergebnissen. Diese beziehen sich auf die Störungsgruppen Affektive Störungen, Angststörungen, Persönlichkeitsstörungen und der Störungen durch psychotrope Substanzen und betreffen den Vorhersagewert von (a) patientenseitigen (b) beziehungsbezogenen und (c) technischen Faktoren (Beutler et al., 2006; Castonguay & Beutler, 2006).

berücksichtigt werden (Clarkin & Levy, 2004). Der *Familienstand* und/oder die *Partnerschaftssituation* wird häufiger im Zusammenhang mit Kontextfaktoren im Sinne einer unterstützenden Beziehung untersucht: Eine bestehende Partnerschaft wird von verschiedenen Autoren als protektiver Faktor bei der Belastung durch psychische Erkrankungen oder der Entwicklung einer Chronifizierung diskutiert (Hahlweg, 2005; Schiepek & Cremers, 2003; Schütz & Wiesner, 2000). Aus den bisherigen Befunden zu Prädiktoren des Behandlungserfolgs lässt sich diese Annahme nicht stützen, da die Partnerschaftssituation in keinem Zusammenhang mit dem Behandlungsergebnis zu stehen scheint (Hölzel et al., 2010; Steffanowski et al., 2007). Dahingegen zeigt sich in dem *sozioökonomischen Status* ein relativ robuster Prädiktor des psychotherapeutischen Behandlungsergebnisses. Dieser wird häufig über eine oder mehrere der folgenden Merkmale operationalisiert: Höchster Schulabschluss, höchster Berufsabschluss, Erwerbsstatus und Einkommen (Clarkin & Levy, 2004). Steffanowski et al. (2007) berichten in ihrer Meta-Analyse, dass ein hoher Schulabschluss sowohl zur Entlassung als auch zur Katamnese mit einem besseren Behandlungsergebnis zusammenhängt. Die Autoren erwägen einerseits einen moderierenden Effekt motivationaler Faktoren, andererseits stellen sie auch die Vermutung an, dass die (in der Rehabilitation psychischer/psychosomatischer Erkrankungen) vorgehaltenen Therapieangebote vor allem Patienten mit höherem Bildungsniveau erreichen. Folglich könnten die therapeutischen Angebote im Hinblick auf das kognitive Anspruchsniveau und die Verständlichkeit überdacht werden. In diesem Zusammenhang liefern Hölzel et al. (2010) ein interessantes Ergebnis: In den Vergleichen zwischen Respondern und Non-Respondern finden sich univariat deutliche Unterschiede im Bildungsniveau. Wurden jedoch weitere Faktoren mit berücksichtigt, zeigte sich, dass dieser Zusammenhang vollständig über die Prozessvariable *Psychotherapie erhalten* vermittelt wurde. Das bedeutet, dass depressive Patienten mit niedrigem Bildungsniveau weniger Psychotherapie (im Rahmen der stationär-psychiatrischen Behandlung) erhalten, die ausbleibenden Effekte also vielmehr auf die nicht vorgehalten Psychotherapie zurückzuführen ist.

Krankheitsbezogene Merkmale

Die krankheitsbezogenen Merkmale *initiale Symptomschwere* und *Chronifizierung* erweisen sich sowohl in der ambulanten Psychotherapie als auch im stationären Bereich als bedeutende Prädiktoren des Outcomes (Barghaan et al., 2005; Beutler et al., 2006; Clarkin & Levy, 2004; Härter et al., 2004; Steffanowski et al., 2007). Die Chronifizierung der Erkrankung wird über die Erkrankungsdauer, aber auch über indirekte Merkmale wie die psychotherapeutische oder psychiatrische ambulante oder stationäre Vorbehandlung oder die Dauer der Krankheitszeiten (Arbeitsunfähigkeitstage) vor Behandlungsbeginn

operationalisiert. Bei Barghaan et al. (2005) erwiesen sich längere Arbeitsunfähigkeitszeiten als relevanter Prädiktor für ein negatives Outcome. Steffanowski et al. (2007) fanden meta-analytisch die Erkrankungsdauer als stärksten Prädiktor für ein schlechteres Therapieergebnis. Die Untersuchung der *Diagnose* als potentiellen Prädiktor des Behandlungsergebnisses fand in den frühen Jahren der Psychotherapieforschung schon eine weite Verbreitung. Kritisch zu betrachten ist jedoch, dass in früheren Untersuchungen nur die erste vergebene Diagnose (bzw. Achse I-Diagnose, s.u.) untersucht wurde, psychische Komorbidität blieb ebenso wie andere Patientenmerkmale, die einen moderierenden Zusammenhang zum Behandlungsergebnis haben könnten, unbeachtet (Clarkin & Levy, 2004). Spätere Studien aus dem angloamerikanischen Raum zum Zusammenhang von psychischer Komorbidität und Behandlungsergebnis zeigten, dass eine komorbide Persönlichkeitsstörung einen relevanten Prädiktor des Outcomes darstellt. Patienten, die neben der symptombezogenen Diagnose, im Sinne einer Achse I Diagnose nach dem DSM IV (Diagnostic and Statistical Manual of Mental Disorders; APA – American Psychiatric Association, 2000; Saß, Wittchen, Houben & Zaudig, 2003), eine Achse II Diagnose einer Persönlichkeitsstörung aufweisen, profitieren weniger von Psychotherapie als Patienten ohne die Diagnose einer komorbiden Persönlichkeitsstörung. Hierzu liegen Ergebnisse für die Erst- bzw. symptombezogenen Diagnosen Depression, Angststörung und Essstörungen vor. Im stationären Setting zeigen sich geringere Behandlungseffekte für somatoforme Störungen im Vergleich zu den anderen Diagnosegruppen. Die besten Behandlungsergebnisse erreichen Patienten mit Anpassungs- und depressiven Störungen (Barghaan et al., 2005). Damit übereinstimmend zeigen auch die meta-analytischen Auswertungen für somatoforme Störungen die geringsten Effektstärken innerhalb der untersuchten Diagnosegruppen (Steffanowski et al., 2007; vgl. auch Abschnitt 2.3.2).

Persönlichkeitsmerkmale

Persönlichkeitsbezogene Merkmale, die im Zusammenhang mit dem Behandlungsergebnis stehen, stellen unter anderem die *Veränderungsbereitschaft* und die *Erfolgserwartung* sowie ein *nichtvermeidender Interaktionsstil* des Patienten dar (Beutler et al., 2006; Clarkin & Levy, 2004). Diese Merkmale gehen sowohl mit einem besseren Therapieergebnis als auch mit geringeren Abbruchraten einher. In einer an die von Beutler et al. (2006) angelehnten Übersichtsarbeit fassen Newman und Stiles (2006) die grundlegenden Befunde zur Vorhersage einer effektiven Behandlung von Angststörungen zusammen. Hinsichtlich der Personenmerkmale finden sich zu den genannten Merkmalen auch eine geringe *internale Kontrollüberzeugung* oder eine hohe *internale Attribution negativer Ereignisse* und eine geringe *Selbstwirksamkeitserwartung* als robuste Prädiktoren für ein negatives Outcome (Newman & Stiles, 2006). Die Merkmale Veränderungsbereitschaft, Erfolgserwartung und

der Hoffnung auf Besserung werden im stationären Setting häufig mittels Instrumenten zur Erfassung der *Psychotherapiemotivation* erhoben. Auch hier zeigt sich der positive Prädiktorwert einer ausgeprägten Psychotherapiemotivation auf die Wirksamkeit von Behandlungen (Barghaan et al., 2005; Nübling, Schulz, Schmidt, Koch & Wittmann, 2006; Schulz, Lang, Nübling & Koch, 2003). Eine ausreichende Psychotherapiemotivation wird auch in der ambulanten Psychotherapieforschung als Prädiktor eines erfolgreichen Therapieergebnisses gesehen (Truant, 1999).

2.4.2 Prozessmerkmale als Prädiktoren des Behandlungserfolgs

Einen Überblick über den Forschungsstand der Prozess-Outcome-Forschung liefern Orlinsky, Ronnestad und Willutzki (2004). Einen relativ robusten Prädiktor des Outcomes stellt dabei die *Güte der therapeutischen Beziehung (therapeutic alliance)* dar (Beutler et al., 2006; Beutler et al., 2004; Martin, Garske & Davis, 2000; Orlinsky et al., 2004), auf die hier näher eingegangen werden soll. Die Untersuchung der differenziellen Wirksamkeit *therapeutischer Verfahren* hat in der Psychotherapieforschung eine lange Forschungstradition und wird hier ebenfalls kurz dargestellt. Der potentielle Einfluss der Prozessmerkmale *Dauer und Dosis* auf das Outcome wird im nächsten Abschnitt gesondert behandelt. Da weitere Faktoren in der vorliegenden Arbeit nicht fokussiert werden und dementsprechend auch nicht untersucht werden sollen, wird auf eine nähere Erläuterung verzichtet. Der Vollständigkeit halber ist auf den Forschungsbereich, der sich mit der Frage nach dem Zusammenhang von weiteren Behandlungs- und Therapeutenmerkmalen mit dem Behandlungsergebnis beschäftigt, hinzuweisen (für einen Überblick vgl. Beutler et al., 2006; Orlinsky et al., 2004).

Therapeutische Beziehung

Die *Güte der therapeutischen Beziehung* stellt einen konsistenten Prädiktor für ein positives Behandlungsergebnis dar (Beutler et al., 2006; Beutler et al., 2004; Martin et al., 2000; Orlinsky et al., 2004). In einer Meta-Analyse über 79 Studien finden Martin et al. (2000) einen signifikanten moderaten Zusammenhang zwischen der therapeutischen Beziehung und dem Behandlungsergebnis (mittlere gewichtete Effektstärke $r = .23$), der relativ durchgängig über die unterschiedlichen Instrumente und Bewertungsperspektiven (Patient, Therapeut oder externer Beobachter) aufzufinden ist. Zu ähnlichen Ergebnissen kommen Beutler et al. (2004) in einer Meta-Analyse über 40 Studien: Die gewichtete mittlere Effektstärke betrug für generische Ergebnisparameter $r = .22$ und $r = .17$ für spezifische Ergebnisparameter. Der angenommene Kausalzusammenhang zwischen therapeutischer Beziehung und Outcome wurde mehrfach in Frage gestellt, da die Beziehung sich im Verlauf der Behandlung verändern kann und eine wechselseitige Beziehung zwischen der therapeutischen

Beziehung und der Symptombesserung möglich ist (Beutler et al., 2004). So kommen auch Tang und DeRubeis (1999b) in einer Untersuchung zu frühen Verbesserungen während kognitiv-behavioraler Depressionsbehandlungen zu dem Schluss, dass eine frühe kognitive Veränderung ein frühes Ansprechen vorhersagt, was wiederum zur Stärkung der therapeutischen Beziehung beiträgt⁹. Allerdings lassen sich auch eine Reihe von Studien finden, in denen eine positive therapeutische Beziehung, auch bei (noch) nicht erfolgter Symptombesserung, einen relativ guten Prädiktor des längerfristigen Outcomes darstellt (für einen Überblick vgl. Beutler et al., 2004). In einer neueren Meta-Analyse von 11 Studien zum Zusammenhang der therapeutischen Beziehung mit dem Therapieabbruch zeigte sich bei einer weniger guten Beziehungsgestaltung eine bedeutend höhere Wahrscheinlichkeit eines Therapieabbruchs (Sharf, Primavera & Diener, 2010).

Insgesamt sind die bisherigen Befunde zum Zusammenhang von frühem Ansprechen und der Güte der therapeutischen Beziehung zum Teil widersprüchlich und zeigen einen bestehenden Forschungsbedarf auf. Zusammenfassend bewerten Beutler et al. (2004) den Prädiktorwert der therapeutischen Beziehung für das Behandlungsergebnis als moderat, wobei die Kausalität des Zusammenhangs nicht gänzlich geklärt ist.

Psychotherapeutische Verfahren

Die Untersuchung der differentiellen Wirksamkeit verschiedener psychotherapeutischer Verfahren bezieht sich hauptsächlich auf die singuläre oder vergleichende Untersuchung von analytisch und tiefenpsychologisch fundierten psychodynamischen Therapien (PDT) und Verhaltenstherapien (VT) bzw. kognitiv-behavioralen Psychotherapien (CBT) (Emmelkamp, 2004; Emmelkamp, Benner, Kuipers, Feiertag, Koster & Van Apeldoorn, 2004; Hollon & Beck, 2004; Leichsenring & Rabung, 2008; Stiles, Barkham, Mellor-Clark & Connell, 2008b). Ein Grund für den Forschungsfokus auf diesen Verfahren, die sich in den theoretischen und anwendungsorientierten Ansätzen stark unterscheiden, liegt in ihrer weiten Verbreitung in der ambulanten und stationären Versorgungspraxis. Eine lange Tradition der differentiellen Therapiewirksamkeitsforschung erbrachte in zahlreichen Meta-Analysen ein Ergebnis, das unter dem Begriff *Dodo-Bird-Verdict*¹⁰ beschrieben wird (Lambert & Ogles, 2004): Demnach erweist sich die Psychotherapie generell zwar als wirksam, die einzelnen Therapieverfahren weisen jedoch keine grundlegenden Unterschiede in der Wirksamkeit auf. In einigen Übersichtsarbeiten zeigt sich eine tendenzielle Überlegenheit der verhaltenstherapeutischen Verfahren, die jedoch häufig auf methodische Artefakte zurückgeführt werden kann (für

⁹ Vgl. auch Abschnitt 2.5.5.

¹⁰ Der Begriff *Dodo-Bird-Verdict* entstammt dem 1865 erschienenen Kinderbuch „*Alice's Adventures in Wonderland*“ des britischen Schriftstellers Lewis Carroll und bezieht sich auf den Ausspruch: „*Everybody has won and all must have prizes*“.

einen Überblick vgl. Lambert & Ogles, 2004). Auch eine neuere Meta-Analyse von Leichsenring und Rabung (2008), die für komplexe psychische Störungen auf eine Überlegenheit von psychodynamischen Langzeittherapien gegenüber kürzeren Therapieverfahren (wie etwa VT) hindeutet, muss unter Berücksichtigung methodischer Einschränkungen, wie etwa die Heterogenität hinsichtlich der Stichproben, der Behandlungsformen, der Kontrollbedingungen und der Ergebnismaße, interpretiert werden (Beck & Bhar, 2009; Kriston, Hölzel & Härter, 2009; Roepke & Renneberg, 2009). Die Frage der differentiellen Wirksamkeit verschiedener Therapieverfahren kann derzeit nicht abschließend geklärt werden.

Neben der Untersuchung der Wirksamkeit unterschiedlicher psychotherapeutischer Verfahren ist hier noch der Forschungsbereich, der sich mit der differentiellen Wirksamkeit von Psychotherapie und Pharmakotherapie sowie deren Kombination auseinandersetzt, zu erwähnen. Einige Studienergebnisse weisen darauf hin, dass (für bestimmte Patientengruppen) eine Kombinationstherapie aus beiden den singulären Verfahren überlegen ist (Cuijpers, van Straten, Warmerdam & Andersson, 2009; Härter et al., 2004; Miller, Norman & Keitner, 1999; Schneider, Härter, Brand, Sitta, Menke, Hammer-Filipiak et al., 2005), wobei in anderen Studien keine Unterscheide zwischen einer pharmako-psychotherapeutischen Kombinationstherapie und alleiniger Psychotherapie gefunden wurden (de Jong-Meyer, Hautzinger, Rudolf, Strauss & Frick, 1996; Teusch, Bohme, Finke, Gastpar & Skerra, 2003). So finden auch Jacobi, Dahme und Rustenbach (1997) in einer Meta-Analyse zur vergleichenden Wirksamkeit von Psycho-, Pharmako- und Kombinationstherapie hinsichtlich der primären Symptomatik und der Depressivität bei Essstörungen (Bulimia und Anorexia nervosa) eine Überlegenheit der Psychotherapie gegenüber der pharmakologischen Behandlung. Die weniger untersuchte Kombinationstherapie erbringt keine höheren Effektstärken als die singuläre psychotherapeutische Behandlung. In Hinblick auf Remissions- und Rückfallraten ist die Kombinationstherapie der reinen Pharmakotherapie überlegen. Die differentielle Wirksamkeit von Psychopharmaka- und Psychotherapie sowie der Kombination beider Verfahren hängt stark von der Erkrankung und anderen Patientenmerkmalen, wie etwa der Krankheitsschwere, sowie von den untersuchten Outcome-Parametern ab (für einen Überblick zur differentiellen Wirksamkeit bei verschiedenen psychischen Störungen vgl. Thase & Jindal, 2004).

2.5 Behandlungsdauer und Symptomverlauf während der Behandlung als Prädiktoren des Behandlungserfolgs

Im letzten Abschnitt wurden empirische Befunde zur Wirksamkeit von Psychotherapie sowie zu Patienten- und Prozessmerkmalen, die mit der Wirksamkeit zusammenhängen, erörtert.

Ein bedeutendes Prozessmerkmal stellt die Behandlungsdauer bzw. genauer die Dosis und Frequenz der Behandlungseinheiten dar. Aufgrund der Fokussierung der vorliegenden Arbeit auf die Untersuchung der Behandlungsdauer und des Symptomverlaufs während der Behandlung und deren Zusammenhang mit dem Behandlungsergebnis werden die theoretischen Konzepte und empirischen Befunde zu den Prozessmerkmalen Behandlungsdauer und Symptomverlauf in diesem Abschnitt gesondert dargestellt.

Die Frage nach dem *Zusammenhang von Dauer/Dosis und Erfolg einer Behandlung* wird im ersten Teil dieses Abschnitts behandelt. Dabei wird zunächst auf die inhaltliche und terminologische Abgrenzung der Begriffe Dauer und Dosis eingegangen (2.5.1). Zur theoretischen Einordnung erfolgt im zweiten Teil eine kurze Übersicht über theoriegeleitete Modelle des Dosis-Wirkungs-Zusammenhangs und des Verlaufs von Psychotherapie (2.5.2). Anschließend werden Ergebnisse aus der ambulanten und der stationären psychotherapeutischen Behandlung zum Dauer-/Dosis-Wirkungs-Zusammenhang dargestellt (2.5.3). Aus der Untersuchung des Dosis-Wirkungs-Zusammenhangs vor allem ambulanter Psychotherapie entwickelte sich eine stärkere Fokussierung auf die Untersuchung des Behandlungserfolgs *im Verlauf der Behandlung*: Neben dem Zusammenhang zwischen der Gesamtdauer und dem kurz- und langfristigen Ergebnis psychotherapeutischer Behandlung stellen die *Prozesse, durch die diese Ergebnisse erreicht werden*, einen Forschungsschwerpunkt in der Psychotherapieforschung dar. Die Verknüpfung von Prozess und Ergebnis findet durch die Untersuchung der Veränderung im Verlauf der Behandlung, z.B. anhand der Symptomatik, statt (Kendall et al., 2004). Aus den Ergebnissen zur Untersuchung des Symptomverlaufs während der Behandlung wurde das Konzept der Early Response, ein frühes Ansprechen auf die Behandlung, entwickelt. Wegen der Relevanz für die vorliegende Arbeit werden spezifische methodische Aspekte der Untersuchung der Early Response und anderer Response-Muster dargestellt (2.5.4). Darauf folgend werden die empirischen Befunde zur frühen Veränderung im Therapieverlauf und deren Zusammenhang zum Therapieergebnis dargestellt (2.5.5), und anschließend wird ein kurzer Überblick über die inhaltlichen Erklärungsansätze des Early Response Phänomens gegeben (2.5.6). Im letzten Teil des Abschnitts wird auf den Einsatz neuer Methoden der Verlaufsanalyse und die bisherigen empirischen Befunde hierzu eingegangen (2.5.7).

2.5.1 Dosis und Dauer

In der ambulanten Psychotherapieforschung wird die Behandlungsdauer häufig äquivalent zur Behandlungsdosis verstanden (vgl. Lambert & Ogles, 2004). In vielen Studien wird der Umfang der Behandlung anhand der einzelnen Sitzungen, also die Dosis, angegeben, ohne dass zusätzlich die Dauer der Behandlung benannt wird. In anderen Studien wird die Wochenanzahl angegeben, ohne im Detail auf die Dosis einzugehen. Grund hierfür ist der

häufig auftretende direkte Zusammenhang von Anzahl der Sitzungen und Wochen, da in der Regel eine Sitzung pro Woche stattfindet. Das lässt sich jedoch keinesfalls verallgemeinern, sodass bei der reinen Wochenangabe etwaige Dosisunterschiede (z.B. wenn die Sitzungen mehrmals pro Woche stattfinden) einen relevanten Unterschied hinsichtlich Dauer-Wirkungs-Zusammenhängen bedingen können (vgl. Ilardi & Craighead, 1994; Tang & DeRubeis, 1999a). Die Integration von Dosis und Dauer, also die an der Dauer relativierte Dosis wird als Behandlungsdichte bezeichnet.

Im stationären Setting stellt sich die Bestimmung der Behandlungsdosis schwieriger dar. Zum einen ist die stationäre Behandlung in Deutschland durch ein multimodales Konzept gekennzeichnet. Hier stellt sich die Frage, auf welche Behandlungsmaßnahmen die Dosisangabe bezogen wird. Für die stationäre Psychotherapie liegt eine Einordnung der Behandlungsmaßnahmen nach psychotherapienahen Verfahren nahe. Barghaan et al. (2005) haben eine detaillierte Einordnung aller in der Rehabilitation psychischer/psychosomatischer Erkrankungen vorgehaltenen Behandlungen in psychotherapienahe Maßnahmen, psychotherapieferne Maßnahmen sowie eine Mittelkategorie vorgenommen. Voraussetzung für eine solche Kategorisierung stellt jedoch die detaillierte Dokumentation der Behandlungselemente dar. Dies ist mit dem Klassifikationssystem KTL (Deutsche Rentenversicherung Bund, 2007) lediglich im Setting der Rehabilitation gegeben. Zudem ist mit der Erfassung und Aufbereitung der Dosis-Angaben ein deutlich erhöhter Aufwand verbunden, was vermutlich auch ein Grund für die seltene Anwendung von Dosisanalysen ist. Einschränkend müssen Dosisanalysen, die über die KTL-Kodierung operationalisiert werden, hinsichtlich der noch ausstehenden Überprüfung der Validität der Prozessangaben mit Vorsicht interpretiert werden (Watzke et al., 2009). Meist wird in Untersuchungen zum Umfang der stationären Behandlung die Dauer als annähernd adäquate Operationalisierung der Dosis bewertet (Steffanowski et al., 2007), da von einem weitgehend proportionalen Verhältnis von Dosis und Dauer ausgegangen wird (Paar & Grohmann, 2000; Schmitz-Buhl, Kriebel & Paar, 1999).

2.5.2 Theoretischer Hintergrund zum Dosis-/Dauer-Wirkungs-Zusammenhang

Das Dosis-Wirkungs-Modell der Psychotherapie nach Howard

Basierend auf einer ersten umfangreichen Metaanalyse aus dem Jahr 1986, die Studien zum Dosis-Wirkungs-Zusammenhang der vorangegangenen 30 Jahre enthielt, entwickelte die Arbeitsgruppe um Howard das *Dosis-Wirkungs-Modell der Psychotherapie* (Howard, Kopta, Krause & Orlinsky, 1986). Nach diesem Modell kommen die bedeutsamsten Veränderungen bereits in einer frühen Phase der Therapie zustande, anschließend kann eine kontinuierliche Verbesserung nur durch eine zunehmend überproportionale Intensität bewirkt werden. An einer Stichprobe von 2 431 Patienten aus 15 Psychotherapiestudien konnten Howard et al.

(1986) zeigen, dass bei der Hälfte der Patienten eine bedeutsame Verbesserung bereits nach der 6. bis 8. Therapiesitzung, bei 75% der Patienten um die 26. Sitzung und ein maximaler Therapieeffekt, d.h. ein Therapieerfolg bei den meisten Patienten, im Durchschnitt nach der 52. Sitzung zu verzeichnen war. Hierbei handelt es sich nicht um einen linearen, sondern um einen logarithmischen Zusammenhang, sodass ein Zuwachs an Besserung mit steigender Therapiedosis geringer wird. Für den Bereich der ambulanten Psychotherapie konnten Hinweise auf die Gültigkeit des Modells gefunden werden (Howard et al., 1986; McNeilly & Howard, 1991).

An den Studien von Howard et al. wird häufig kritisiert, dass sie in der Regel auf Prä-Post-Untersuchungen mit lediglich zwei Messzeitpunkten basieren, katamnestische Einschätzungen fehlen ebenso wie eine detaillierte Erfassung der Ergebniskriterien im Verlauf der Behandlung (Paar & Grohmann, 2000). Problematisch ist ferner die undifferenzierte Betrachtung der Ergebniskriterien. Nach Strauß und Wittmann (2005) ist davon auszugehen, dass für unterschiedliche Ziele in der Psychotherapie auch unterschiedlicher Aufwand benötigt wird. Der benötigte zeitliche Aufwand sollte demnach differenziert für einzelne Ergebniskriterien untersucht werden.

Weiterentwicklung zum Phasenmodell psychotherapeutischer Veränderungen

Theoriegeleitet und empirisch konnte gezeigt werden, dass sich unterschiedliche Bereiche im Erleben und Verhalten des Patienten im Verlauf der Psychotherapie in unterschiedlichem Ausmaß verändern (Howard, Lueger, Maling & Martinovic, 1993; Lueger, 1995). In dem nach Howard et al. (1993) entwickelten Phasenmodell werden drei Phasen der psychotherapeutischen Veränderung unterschieden:

- Wiederherstellung des Wohlbefindens (Remoralisierung)
- Symptomverbesserung (Remediation)
- Verbesserung des Funktionsniveaus in verschiedenen Lebensbereichen (Rehabilitation)

In der ersten Phase wird in einer relativ kurzen Zeit eine Verbesserung des subjektiv erlebten Wohlbefindens erreicht. Dies scheint durch unspezifische Faktoren, wie das Gefühl, mit der Erkrankung akzeptiert zu werden und wieder Hoffnung zu schöpfen, zu erfolgen. Die erste Phase stellt eine Voraussetzung für die Veränderung in der Symptombelastung dar. In der zweiten Phase kommt es zu einer gerichteten, systematischen Reduktion der Krankheitssymptome, was durch das Erkennen und Bewerten der Symptome, die Aktivierung von vorhandenen Ressourcen und das Erlernen alternativer adaptiver Bewältigungsstrategien erreicht wird. In der dritten Phase der Veränderung wird die

Funktionsfähigkeit in unterschiedlichen relevanten Lebensbereichen fokussiert. Stabile Langzeiteffekte psychotherapeutischer Behandlungen resultieren aus dem optimalen Durchlaufen aller drei Therapiephasen (Paar & Grohmann, 2000).

Anhand des Phasenmodells therapeutischer Veränderungen können die einzelnen Ergebnisbereiche empirisch überprüft werden, da diese Bereiche relativ valide zu erfassen sind. Mittlerweile konnte das Phasenmodell in einigen Studien empirisch bestätigt werden (z.B. Barkham, Rees, Stiles, Shapiro, Hardy & Reynolds, 1996; Lueger, 1998). Nach Paar und Grohmann (2000) lässt sich das Modell auch auf die psychosomatische Rehabilitation übertragen.

2.5.3 Empirische Befunde zum Dosis-Wirkungs-Zusammenhang und Symptomverlauf während der Behandlung

Die Frage einer angemessenen Therapiedauer und -dosis wurde im englischsprachigen Raum hauptsächlich im Bereich der *ambulanten Psychotherapieforschung* untersucht (Haas, Hill, Lambert & Morrell, 2002; Howard et al., 1986; Kelly, Roberts & Ciesla, 2005; Lambert, Hansen & Finch, 2001; Lutz, Bachmann, Tschitsaz, Smart & Lambert, 2007a; Lutz, Rafaeli, Howard & Martinovich, 2002; McNeilly & Howard, 1991; Stiles, Barkham, Iveson, Iveson, Leach, Lucock et al., 2003; Tang & DeRubeis, 1999b; Thompson, Thompson & Gallagher-Thompson, 1995). Ein Überblick über die hierzu durchgeführten Metaanalysen ist bei Lambert und Ogles (2004) zu finden. Die von Howard et al. (1986) postulierte logarithmische Beziehung zwischen der Anzahl der Stunden und der Therapieerfolgsrate konnte empirisch mehrfach bestätigt werden (für einen Überblick vgl. Howard et al., 1986; McNeilly & Howard, 1991). Einschränkend ist jedoch für viele Studien die Eindimensionalität des Ergebniskriteriums sowie das Prä-Post-Design zu nennen (vgl. auch Abschnitt 2.2). Spätere Studien versuchten der oben genannten Kritik Rechnung zu tragen und untersuchten den Zusammenhang zwischen Dosis und Wirkung, indem sie die Datenerhebung nach jeder Therapiesitzung durchführten (Anderson & Lambert, 2001; Kadera, Lambert & Andrews, 1996). Anderson und Lambert (2001) konnten in der ambulanten Psychotherapie nach der 13. Sitzung eine klinisch signifikante Verbesserung bei 50% der Patienten erfassen, nach der 50. Sitzung traf dies auf 75% der Patienten zu. In neueren theoretisch-konzeptuellen Ansätzen aus der ambulanten Psychotherapieforschung werden Dauer-Wirkungs-Zusammenhänge als Ergebnis einer gemeinsamen Entscheidungsfindung zwischen Behandler und Patienten verstanden. Anders als im klassischen Dosis-Wirkungs-Modell von Howard et al., in dem der Frage nach einer angemessenen bzw. ausreichenden Therapiedosis nachgegangen wird, steht hier die Einbeziehung von Kontextfaktoren bei der Bestimmung der Behandlungsdauer im Vordergrund. In dem Konzept des *responsive regulation model* wird postuliert, dass die Dauer einer Behandlung anhand vieler

Kontextfaktoren festgelegt wird, die Zielerreichung im Sinne einer Symptombesserung im Prozess gemeinsam überprüft und die Therapiedauer im Verlauf entsprechend angepasst werden kann (Stiles, Barkham, Connell & Mellor-Clark, 2008a; Stiles, Honos-Webb & Surko, 1998). Faktoren, die zur Festlegung der Behandlungsdauer beitragen, können ressourcenbasierte Merkmale, Setting- und Therapeutenmerkmale, vor allem aber patientenseitige Charakteristika sein. Berücksichtigt werden müssen auch entsprechende Unterschiede in den psychotherapeutischen Verfahren, die sich neben theoretisch-konzeptionellen Inhalten auch hinsichtlich des Umfangs unterscheiden. So zeigt die Meta-Analyse von Leichsenring und Rabung (2008), die auf eine tendenzielle Überlegenheit psychodynamischer Langzeittherapien gegenüber kürzerer Therapieformen hinweist, die schwer trennbare Verknüpfung von Behandlungsverfahren und Behandlungsdauer auf. Ebenso muss die Passung von Patient und Therapeut, im Sinne der Güte der therapeutischen Beziehung, bzw. von Patient und Therapieverfahren, im Sinne einer differentiellen Indikationsstellung, berücksichtigt werden. Watzke, Rüdell, Jürgensen, Koch, Kriston, Grothgar et al. (2010) verglichen die Wirksamkeit von stationärer Psychotherapie bei systematischer und zufälliger Zuweisung zu verhaltenstherapeutischen und psychodynamischen Behandlungsverfahren und fanden nur bei den psychodynamischen Verfahren eine Überlegenheit der systematischen Zuweisung. Insgesamt wird deutlich, dass der Zusammenhang zwischen Dauer und Wirksamkeit nicht unabhängig von den Kontextfaktoren untersucht werden kann.

Für die *stationären psychotherapeutischen Behandlung* liegen bislang wenige Studien zum Dosis-Wirkungs-Zusammenhang vor (Neeb et al., 2001; Nosper, 1999a; Schmitz-Buhl et al., 1999; Zielke, 2006). Einige Studien weisen dabei auf einen positiven Zusammenhang zwischen Therapiedauer und Wirksamkeit hin (Borgart & Meermann, 1999; Hartmann & Zepf, 2004; Nosper, 1999a). Ebenso zeigte eine Analyse im Kontext der Bildung von Fallgruppen in der Behandlung von psychischen Störungen, dass die Patienten mit der längsten Behandlungsdauer auch die besten Behandlungsergebnisse erzielten und diejenigen mit der kürzesten Behandlungsdauer sich am wenigsten verbesserten (Andreas et al., 2009; Andreas, Dirmaier, Koch & Schulz, 2005). Hinsichtlich langfristiger Effekte unterschiedlicher Behandlungsdauern konnten Lamprecht und Schmidt (1990) in der sogenannten Zauberberg-Studie zeigen, dass eine mittlere Behandlungsdauer von 6 bis 8 Wochen gegenüber längeren (mehr als 8 Wochen) und kürzeren (4 bis 6 Wochen) Behandlungen bessere Ergebnisse in verschiedenen patientenseitig beurteilten Kriterien am Ende der Rehabilitation sowie katamnestisch nach einem und nach drei Jahren erbrachte. In einer Arbeit von Nosper (1999a) wurde der Erfolg der psychosomatischen Rehabilitationsbehandlung in Abhängigkeit von der Behandlungsdauer an einer Stichprobe

von 266 stationär behandelten Patienten katamnestisch untersucht. Die Outcome-Maße waren dabei die Allgemeine Depressionsskala (ADS-L; Hautzinger & Bailer, 1993) und die SCL-90-R (Franke, 2002). Nach der Behandlungsdauer von 3 Wochen wurden zwar schon signifikante Verbesserungen der Symptomreduktion mit kleiner bis mittlerer Effektstärke ($d = 0.40$) nachgewiesen. Nach sechswöchiger Behandlung konnten die Effektstärken auf Werte zwischen $d = 0.65$ und $d = 0.80$ gesteigert werden, die auch katamnestisch nach einem Jahr noch in mittlerer Ausprägung nachweisbar waren ($d = 0.45$ bis $d = 0.65$). Nosper schlussfolgert, dass erst bei einer Behandlungsdauer von mehr als 3 Wochen klinisch bedeutsame und langzeitstabile Behandlungseffekte erreicht werden (Nosper, 1999a). Damit übereinstimmend zeigen die Ergebnisse von Schmitz-Buhl et al. (1999), dass eine bedeutsame Steigerung der Wirksamkeit erst nach einer Behandlungsdauer von 3 bzw. 4 Wochen eintritt. Zielke (2006) entwickelte auf der Basis einer Jahresstichprobe aus dem Jahr 2003 von Patienten aus der stationären verhaltensmedizinischen Rehabilitation sogenannte differentielle Behandlungszeitkorridore, die eine annähernd individualisierte Behandlungsdauer für jeden Patienten in Abhängigkeit von den vier empirisch begründeten Einflussgrößen Lebensalter, Behandlungsdiagnose, Komorbidität und Krankheitsdauer eingrenzen. Steffanowski et al. (2007) fanden in der Meta-Analyse zur Wirksamkeit der stationären psychosomatischen Rehabilitationsbehandlung in der Behandlungsdauer einen bedeutenden positiven Prädiktor für das Behandlungsergebnis. Dieser Zusammenhang fiel katamnestisch sogar höher aus als zum Entlasszeitpunkt. In der Übersichtsarbeit zur Wirksamkeit von stationärer Gruppenpsychotherapie von Kösters et al. (2006) konnten keine Zusammenhänge der Behandlungsdauer oder der Dosis (Anzahl der Sitzungen) mit dem Behandlungsergebnis nachgewiesen werden. Ebenso wurde in einer Untersuchung der Therapieresponse bzw. - Non-Response in der stationären Depressionsbehandlung kein Zusammenhang von Behandlungsdauer und Therapieresponse gefunden (Hölzel et al., 2010).

Aus internationalen Studien, deren Ergebnisse aufgrund der Unterschiede der Settings im Vergleich zur stationären Versorgung in Deutschland nur eingeschränkt übertragbar sind, resultieren ebenfalls uneinheitliche Ergebnisse: In einer Untersuchung der stationären Behandlung depressiver Patienten zeigte sich, dass längere Behandlungen mit einer Reduktion der depressiven Symptomatik und einer Verbesserung des allgemeinen Funktionsniveaus einhergehen, in anderen Outcome-Bereichen aber keine Unterschiede zwischen länger und kürzer behandelten depressiven Patienten auftraten (Lieberman, Wiitala, Elliott, McCormick & Goyette, 1998). Aus den Ergebnissen einer Untersuchung der Behandlungsdauer von 458 Patienten mit unipolarer Depression (Lauber, Lay & Rössler, 2006) schlussfolgern die Autoren, dass ein nicht-linearer Zusammenhang in Form einer umgekehrt U-förmigen Kurve zwischen Behandlungsdauer und -wirksamkeit besteht:

Patienten die weniger als 15 Tage behandelt wurden zeigten die geringste Verbesserung, eine Behandlungsdauer von 61 bis 100 Tagen ging mit der größten Verbesserung einher, wohingegen längere Behandlungen wiederum zu geringeren Verbesserungsraten führen.

Die zum Teil widersprüchlichen Ergebnisse zum Dauer-Wirkungs-Zusammenhang im stationären Setting können zum einen auf die Heterogenität der Versorgungsstrukturen bzw. der behandelten Patientenstichproben zurück geführt werden. Andererseits muss auch der zeitliche Kontext, in dem die Untersuchungen durchgeführt wurden, berücksichtigt werden: Aufgrund der beobachtbaren Tendenz der Verkürzung und Vereinheitlichung der Behandlungszeiten (vgl. Abschnitt 2.1.2) beziehen sich ältere Studien auf Behandlungsmodalitäten, die zu den aktuellen Bedingungen divergent sein können. Die Verkürzung der Behandlungszeiten ist sowohl für das deutsche Versorgungssetting (z.B. Steffanowski et al., 2007) als auch im angloamerikanischen Bereich zu beobachten. So zeigen beispielsweise Lieberman et al. schon 1996 deutliche Verkürzungen der Behandlungsdauer in den vorangehenden acht Jahren auf (Lieberman et al., 1998). Letztlich stellt sich die Untersuchung des Zusammenhangs von Behandlungsdauer und Behandlungsergebnis unter naturalistischen Bedingungen bei einer eingeschränkten Varianz der Dauer erschwert dar. So können in der Rehabilitation psychischer/psychosomatischer Erkrankungen nur sehr eingeschränkte Behandlungszeiten vergleichend untersucht werden, da die Mehrheit der Patienten eine Behandlungsdauer zwischen 4 und 6 Wochen erhält (Barghaan et al., 2005; Deutsche Rentenversicherung Bund, 2010). Zudem muss bei naturalistischen Studien berücksichtigt werden, dass eine indikationsspezifische Anpassung der Behandlungsdauer, sodass die Patienten je nach Bedarf und Erkrankungsschwere adäquat lang behandelt werden, einen gering oder gar nicht ausgeprägten Zusammenhang von Dauer und Ergebnis der Behandlung zur Folge haben könnte. Die oben erläuterte Verknüpfung zwischen Behandlungsdauer und Therapieverfahren kommt im stationären Setting weniger stark zum Tragen, da die Dauer hier stärker an institutionelle bzw. kostenträgerbezogene Vorgaben und weniger an therapieschulenspezifische Konzeptionen gebunden ist (vgl. Abschnitt 2.1.2).

Einschränkend ist darauf hinzuweisen, dass aufgrund des durchgängigen Fehlens experimenteller Studiendesigns, die bei dieser Art von Studien aus klinisch-ethischen Gründen nur sehr schwer realisierbar sind, eindeutige kausale Schlussfolgerungen nicht möglich sind. Insbesondere ist hierbei zu beachten, dass Patienten mit unterschiedlichen Behandlungsdauern in der Regel auch in anderen Variablen (z.B. Ausgangsbelastung) systematische Unterschiede aufweisen können (vgl. Abschnitt 2.2). Diese potentiellen Unterschiede können in nicht-experimentellen Designs im besten Fall statistisch kontrolliert

werden und müssen bei der Interpretation der Ergebnisse berücksichtigt werden. Eine experimentelle Zuweisung der Patienten zu bestimmten Behandlungszeiten würde zu einem besseren Verständnis der Kausalbeziehung zwischen Behandlungsdauer und -ergebnis führen, ist aber aus ethischen Gründen nicht realisierbar. Dabei ist die Frage nach Dauer/Dosis-Wirkungs-Zusammenhängen in der stationären Versorgung auch unter gesellschaftspolitischen Perspektiven, vor allem in Hinblick auf die Diskussion um Verkürzungen der Behandlungsdauer von großer Relevanz. Anhand des derzeitigen Forschungsstandes lässt sich keine ausreichende Evidenz für adäquate, effektive und effiziente Behandlungszeiten ableiten.

Eine Erweiterung der Untersuchungen des Dauer-Wirkungs-Zusammenhangs liefern die zum Teil oben schon erwähnten Studien, die die Symptombelastung kontinuierlich während der Behandlung erfassen. Mithilfe der Verlaufsuntersuchung ist es möglich, ein differenziertes Abbild der Veränderungen im Behandlungsverlauf, Erkenntnisse über den Zeitpunkt, zu dem Verbesserungen oder Verschlechterungen auftreten, sowie zu Mustern der Stabilität oder Diskontinuität zu erhalten.

Im Zuge der Verlaufsuntersuchung psychotherapeutischer Behandlung ließen sich interindividuell unterschiedliche Verläufe finden, was zu einer Fokussierung auf die Identifikation von bestimmten Patientensubgruppen mit distinkten Symptomverläufen führte. Stehen diese Verlaufsgruppen in einem Zusammenhang mit dem langfristigen Behandlungserfolg, so kann die Kenntnis von distinkten Symptomverläufen zu einer individualisierten, patientenorientierten Optimierung der Behandlung beitragen. Es können Patienten mit positiven aber auch mit negativen Therapieverläufen frühzeitig in der Behandlung erkannt und entsprechende Modifikationen der Therapie vorgenommen werden. Die Untersuchung von distinkten Symptomverläufen während der psychotherapeutischen Behandlung nimmt in der vorliegenden Arbeit eine zentrale Rolle ein. Aus diesem Grund wird im Folgenden spezifisch auf die Methoden und Ergebnisse der Studien zur Untersuchung distinkter Symptomverläufe eingegangen. Das theoretisch und empirisch am besten belegte Verlaufsmuster beschreibt die *Early Response*. Nach der Darstellung und der methodischen Auseinandersetzung mit dem Konzept werden abschließend neue methodische Entwicklungen und Ergebnisse der Untersuchung des Behandlungsverlaufs aufgezeigt.

2.5.4 Das Konzept der Early Response

Die folgende Darstellung der Konzeptbildung und des aktuellen Forschungsstands zum Phänomen der Early Response orientiert sich hauptsächlich an den Übersichtsarbeiten von Lambert und Lambert und Ogles (Lambert, 2005; Lambert & Ogles, 2004).

Der Begriff der *Early Response* stammt aus der pharmakologischen Forschung, in der eine *Response*, also ein Ansprechen auf eine Medikation, über den zeitlichen Eintritt der Wirksamkeit definiert wird. Eine sogenannte *Early Response* beschreibt dort ein verfrühtes Auftreten der Wirksamkeit, das nicht auf die physiologische Wirkung der Substanz zurückgeführt werden kann, und daher auch als eine *Placebo Response* bezeichnet wurde. Frühere pharmakologische Studien wiesen auf einen Zusammenhang zwischen früher (Placebo-) Response und einem schlechteren mittel- und langfristigen Behandlungsergebnis – im Vergleich zu einer verzögerten Response – hin (Quitkin, McGrath, Stewart, Taylor & Klein, 1996; Stewart, Quitkin, McGrath, Amsterdam, Fava, Fawcett et al., 1998). Neuere Studien zum zeitlichen Wirkeintritt von Antidepressiva weisen jedoch vermehrt auf einen positiven prädiktiven Wert des frühen Ansprechens für das Behandlungsergebnis hin (Katz, Tekell, Bowden, Brannan, Houston, Berman et al., 2004; van Calker, Zobel, Dykieriek, Deimel, Kech, Lieb et al., 2009).

In der Psychotherapieforschung wurde der Zeitpunkt des Ansprechens während der Behandlung erst in den letzten 20 Jahren stärker beforscht. Hier finden sich starke Hinweise für ein ähnliches Bild: Patienten, die im Laufe der Behandlung früh auf diese ansprechen, haben sowohl am Ende der Behandlung als auch zu Follow-up-Messungen ein besseres Behandlungsergebnis als Patienten, die verzögert oder gar nicht respondieren (Ilardi & Craighead, 1994; Lambert, 2005; Renaud, Brent, Baugher, Birmaher, Kolko & Bridge, 1998; Tang & DeRubeis, 1999b). Allerdings gibt es derzeit keine einheitliche Definition der Early Response oder anderer Verlaufsbeschreibungen. Die Early Response (ebenso wie die Non-Response oder andere potentielle Verläufe) kann durch mehrere Parameter bestimmt werden (Lambert, 2005). Zum einen stellt sich die Frage nach dem qualitativen Kriterium, nämlich welches Konstrukt sich in welchem Ausmaß (nicht) verändert haben muss, um von einer (Non-)Response zu sprechen. Das am häufigsten untersuchte Konstrukt ist die Symptombelastung. Begründet ist dies darin, dass dieses Konstrukt auch einen Outcome-Parameter darstellt, der sehr häufig in Wirksamkeitsstudien untersucht wird (vgl. Abschnitt 2.3.2). Zur Untersuchung des Response-Zeitpunktes während der Behandlung werden die gleichen Maße eingesetzt wie zur Untersuchung des Behandlungsergebnisses in klassischen Outcome-Studien. Hinsichtlich des Kriteriums des Ausmaßes des Ansprechens wird eine Vielzahl von Operationalisierungen verwendet. Generell kann zwischen absoluten und relativen Maßen unterschieden werden. Als absolutes Maß gilt der ausschließliche Zustand, z.B. das Ausmaß der Symptombelastung, zu einem definierten Zeitpunkt. Relative Maße bilden hingegen die Veränderung in Abhängigkeit zum Ausgangswert, z.B. über Differenzwerte zwischen Ausgangswert und dem Wert zu einem kritischen Zeitpunkt oder über die prozentuale Abnahme der Symptomatik. Einige Autoren wählen als absolutes Maß das Kriterium einer über den Therapeuten beurteilten vollständigen oder beinahe

vollständigen Remission, also ein vollständiges Ausbleiben der Psychopathologie, unabhängig von der Ausgangsbelastung (Stewart et al., 1998). Dieses Maß ist durch die unstandardisierte Beurteilung des Therapeuten und die sehr konservative Definition einer Response als vollständige „Heilung“, die z.B. bei chronifizierten Patienten nur schwer erreichbar ist, wenig nützlich, um Response zu operationalisieren. Ein Kriterium, das die Vergleichbarkeit von Studienergebnissen ermöglicht, ist das Unterschreiten eines Cut-Off-Wertes in einem standardisierten Outcome-Instrument (Hautzinger, de Jong-Meyer, Treiber, Rudolf, Thien & Bailer, 1996b), welcher aus Referenzstichproben hergeleitet werden kann. In anderen Studien werden relative Veränderungsmaße, wie die prozentuale Verbesserung z.B. um 50% (Beckham, 1989; Renaud et al., 1998; van Calker et al., 2009) oder um 20% bzw. 25% (Tadic et al., 2010; van Calker et al., 2009; Van, Schoevers, Kool, Hendriksen, Peen & Dekker, 2008), das Erreichen einer Effektstärke mit einer bestimmten Größe (Lutz et al., 2009) oder das Überschreiten einer kritischen Differenz¹¹ (Teil des Kriteriums bei Lutz et al., 2007a; Tang & DeRubeis, 1999b) zu einem a priori festgelegten Zeitpunkt im Vergleich zum Ausgangswert, angewendet.

Neben dem gewählten Konstrukt und dem definierten Ausmaß der Veränderung stellt das zeitliche Kriterium den dritten Parameter zur Bestimmung der (Non-)Response dar. Zu welchem Zeitpunkt nach Behandlungsbeginn von einer früher Response gesprochen wird, variiert in den Studien stark, von wenigen Wochen (z.B. 2 Wochen bei Tadic et al., 2010) bis zu mehreren Monaten nach Beginn der Behandlung (z.B. 2 Monate bei Van et al., 2008). Häufig hängt die Festlegung des Zeitpunkts einer Early Response von der Gesamtdauer der Psychotherapie ab. Diese wiederum unterscheidet sich alleine aufgrund theoretisch-konzeptioneller Grundlagen der unterschiedlichen Therapieschulen. Die Behandlungsdauer bei depressiven Erkrankungen stellt sich bei kognitiver Verhaltenstherapie (Ilardi & Craighead, 1994) erwartungsgemäß kürzer dar als bei psychodynamischen Verfahren (Van et al., 2008).

Entsprechend der uneinheitlichen Verwendung der Beschreibung eines frühen Ansprechens auf eine Behandlung ist das verzögerte oder späte Ansprechen uneinheitlich definiert. Die Begriffe *Delayed* oder *Late Response* gehen ebenso wie die *Early Response* auf die pharmakologische Forschung zurück. Dabei wird das Kriterium der Latenzzeit ebenso wie bei der Early Response häufig stichprobenabhängig festgelegt (z.B. Kaplan, Lazoff, Kelly, Lukin & Garver, 1990). In einer Studie von Van, Dekker, Koelen, Kool, van Aalst, Hendriksen et al. (2009) wurden depressive Patienten, die nach den ersten 8 Wochen einer psychodynamischen Kurzzeittherapie (16 Sitzungen in 24 Wochen) keine ausreichende Verbesserung zeigten, d.h. eine mindestens 50%ige Verbesserung in der HAMD (Hamilton,

¹¹ Vgl. Abschnitt 2.3.1.

1960), im weiteren Verlauf bis zur 24. Woche jedoch das Kriterium einer Response erreichten, als *Delayed Response* bezeichnet. Ähnlich erfolgt die Einteilung in Response-Gruppen (> 50% Verbesserung in der HAM-D-17) in einer Studie zur Wirklatenz von Antidepressiva (Trivedi, Morris, Grannemann & Mahadi, 2005): Patienten, die innerhalb der ersten 4 Wochen auf die Behandlung ansprachen, wurden als *Early Response* Gruppe bezeichnet, eine *Late Response* war bei den Patienten gegeben, die zwischen der 4. und 12. Woche respondierten. Wurde das Kriterium nach mehr als 12 Wochen nicht erreicht, wurde eine Non-Response definiert.

Klarer beschrieben ist die *Non-Response*, wenn Patienten auf eine Therapie überhaupt nicht bzw. nicht im vorher definierten Ausmaß ansprechen. Diese wird zeitlich meist zum Ende der Behandlung festgelegt, das inhaltliche Kriterium, also Ausmaß der Veränderung, das hier *nicht* überschritten wird, variiert in Abhängigkeit von der Definition der Response.

Letztlich trägt die Vielfältigkeit in der Operationalisierung des Behandlungsergebnisses zu der uneinheitlichen Definition und damit auch zu uneinheitlichen Ergebnissen in der Untersuchung von Behandlungsverlauf und -ergebnis bei (Tedlow, Fava, Uebelacker, Nierenberg, Alpert & Rosenbaum, 1998). Hier stellen sich die gleichen Fragen für die Response im Verlauf: Welches Konstrukt wird als Ergebnisparameter definiert, über welches Instrument wird das Ergebnis operationalisiert (das gleiche wie in der Verlaufsmessung oder ein anderes), nach welchem Kriterium wird eine ausreichende Verbesserung bemessen (Response oder Remission) und zu welchem Zeitpunkt wird es gemessen (zum Ende der Behandlung und/oder zu einem Katamnese-Zeitpunkt)?

2.5.5 Early Response als Prädiktor des Behandlungserfolgs

In einer Übersichtsarbeit aus dem Jahr 1994 wurden acht Studien aus dem Zeitraum von 1980 bis 1993 zu den Wirkfaktoren der kognitiv-behavioralen Therapie (Cognitive-Behavior Therapy; CBT) bei depressiven Erkrankungen zusammengefasst (Ilardi & Craighead, 1994). Es zeigte sich, dass 60%-70% der durchschnittlichen Symptomreduktion bei Behandlungsende bereits innerhalb der ersten vier Behandlungswochen auftraten. Die Autoren interpretieren diese frühen Verbesserungen im Sinne eines Ansprechens auf unspezifische Wirkfaktoren wie etwa der Veränderungsbereitschaft (s.u.). Diese Schlussfolgerungen führten zu einer kontroversen Diskussion. Tang und DeRubeis (1999a) kritisierten dabei, dass der Aspekt, dass die CBT in den ersten 4 Wochen in einer relativ hohen Psychotherapiedosis (zwei mal wöchentlich) durchgeführt wird, unberücksichtigt blieb. Haas et al. (2002) entwickelten eine Methode zur Identifikation von Early Response-Patienten, die nicht stichprobenabhängig ist, jedoch auch nicht von einer vollständigen Remission ausgeht. Anhand eines großen Referenzdatensatzes von über 10 000 Personen mit Messungen nach jeder Sitzung bildeten die Autoren mittels hierarchisch linearer

Regression erwartete Verbesserungskurven für Patientengruppen mit unterschiedlich stark ausgeprägter Belastung zu Behandlungsbeginn. Anhand der Untersuchungsstichprobe von 147 Personen konnten so die Abweichungen der beobachteten Verläufe jedes einzelnen Patienten von der Referenzgruppe der erwarteten Verläufe ermittelt und die Early Response über ein starke, frühe Abweichung des erwarteten Verlaufs definiert werden. Als qualitatives Kriterium einer Verbesserung wendeten die Autoren das von Jacobson und Truax eingeführte Kriterium der klinischen Signifikanz an (Jacobson & Truax, 1991). Zur Untersuchung des Vorhersagewertes des Behandlungsverlaufs auf das Ergebnis wurden hierarchische Regressionen mit der Ausgangsbelastung als zu kontrollierender erster Prädiktor durchgeführt. Sowohl zum Ende der Behandlung als auch zur Follow-up-Messung konnte durch den Behandlungsverlauf, auch bei Kontrolle der Symptombelastung zu Behandlungsbeginn, ein essentieller Anteil der Varianz im Behandlungsergebnis aufgeklärt werden (inkrementelles $R^2_{\text{Entlassung}} = .40$; inkrementelles $R^2_{\text{Follow-up}} = .19$; beide auf dem 1%-Niveau signifikant). Bisherige Ergebnisse zum positiven Zusammenhang von Early Response und langfristiger Symptombesserung konnten auch mit diesem differenzierten methodischen Vorgehen bestätigt werden (Haas et al., 2002).

Zur Beschreibung von diskontinuierlichen Verläufen führten Tang und DeRubeis (1999b) den Begriff der *Sudden Gains* ein, der die plötzliche Symptomverbesserung zwischen zwei aufeinander folgenden Sitzungen sowie die Stabilität vor und nach der Veränderung beschreibt. Das Kriterium einer plötzlichen relevanten Verbesserung setzt sich aus einem Algorithmus zusammen, in dem die Verbesserung in einem bestimmten Ausmaß ins Verhältnis zu den vorangegangenen und den folgenden Sitzungen gestellt wird. Mehrere Untersuchungen konnten zeigen, dass Sudden Gains früh im Behandlungsverlauf auftreten und besonders rasche Verbesserungen (*Early Sudden Gains*) mit dem größten Therapieerfolg einhergehen (Kelly et al., 2005; Lutz et al., 2007a; Stiles et al., 2003). In einer Replikation ihrer Studie konnten Tang, DeRubeis, Hollon, Amsterdam und Shelton (2007) zeigen, dass bei den Patienten, bei denen während der Behandlung Sudden Gains der Symptomatik auftraten, es auch langfristig zu weniger Rezidiven der Erkrankung kam. In einer Untersuchung von Sudden Gains in der *teilstationären* Behandlung depressiver Patienten ($n = 664$, mit einer mittleren Behandlungsdauer von 33.6 Tage und einer Standardabweichung von 15.8 Tagen) wurde die depressive Symptomatik in wöchentlichen Intervallen erhoben (Drymalski & Washburn, 2011): Sudden Gains wurden hier als Veränderung zwischen zwei Messungen in dem Mindestausmaß einer *reliable change*¹² sowie einer mindestens 25%igen Verbesserung zum Ausgangswert definiert. Es zeigte sich, dass bei 41% der Patienten Sudden Gains auftraten. Von diesen 41% wiesen fast alle

¹² Vgl. Abschnitt 2.3.1.

Patienten (93%) ausschließlich *einen* Verbesserungssprung auf, der wiederum bei dem größten Teil der Gruppe (78%) innerhalb der ersten 2 Behandlungswochen auftrat. Hinsichtlich des Behandlungsergebnisses zeigte sich, dass die Gruppe der Patienten, bei denen Sudden Gains auftraten, ein bedeutend besseres Outcome (Symptomschwere und Lebensqualität) zum Ende der Behandlung aufwiesen als die Patientengruppe ohne Sudden Gains. Zusammenhänge zum längerfristigen Outcome wurden jedoch nicht erhoben (Drymalski & Washburn, 2011). Die Ergebnisse zum Konzept der Sudden Gains stützen das Konzept der Early Response. Einen Überblick über die bis zum Jahr 2005 veröffentlichten Ergebnisse zur Early Response in der ambulanten Psychotherapieforschung ist bei Lambert (2005) zu finden.

2.5.6 Inhaltliche Erklärungsansätze der Early Response

Die Frage, welche Wirkfaktoren bei einer Early Response zum Tragen kommen, wird kontrovers diskutiert. Illardi und Craighead erklären die Befunde der hohen Symptomreduktion in der frühen Phase der Behandlung mit CBT durch unspezifische Wirkfaktoren, wie die Hoffnung und Bereitschaft zur Veränderung, die hier zu einer frühen Response und langfristigen Wirksamkeit beitragen (Illardi & Craighead, 1994). Ähnliche Ergebnisse aus Untersuchungen bei anderen Indikationen stützen die Annahme von unspezifischen Wirkfaktoren (Wilson, 1999). Dagegen sehen Tang und DeRubeis das Phänomen der Early Response in den ersten Wochen der kognitiven Verhaltenstherapie in spezifischen therapeutischen Techniken, die in der ersten Phase hoch dosiert (zwei Sitzungen pro Woche) zum Einsatz kommen, begründet (Tang & DeRubeis, 1999b). Bisherige Studien zu den Wirkmechanismen der Early Response liefern zwar Hinweise auf allgemeine Wirkfaktoren der Psychotherapie, es bedarf jedoch weiterer Untersuchungen, um die Frage nach spezifischen und unspezifischen Wirkfaktoren zu beantworten.

2.5.7 Neuere Methoden der Untersuchung von Symptomverlaufsgruppen während der Behandlung

In den letzten Jahren kommen zunehmend neue, explorative Methoden zur Identifikation von Verlaufs- oder Response-Gruppen zum Einsatz. Diese Methoden ermöglichen die Analyse aller Messzeitpunkte, ohne an die Restriktionen der klassischen Verfahren, wie Varianzanalysen mit Messwiederholung oder Zeitreihenanalysen, gebunden zu sein. Zwei Verfahren stehen dabei im Vordergrund: Bei der *Hierarchischen Linearen Modellierung* (HLM; Bryk & Raudenbusch, 1992) werden die Wachstumskurven jedes Patienten geschätzt und der Gesamtverlauf in Abhängigkeit von Patientenmerkmalen modelliert; die Methode des *Growth Mixture Modeling* (GMM; Eid, Geiser & Nußbeck, 2008; Jung & Wickrama, 2008) stellt eine Erweiterung des HLM-Ansatzes dar, bei dem nicht von *einer* Population mit einem

bestimmten Verlauf ausgegangen wird, sondern von Subgruppen mit distinkten Verläufen, die mittels GMM identifiziert werden können (für eine detaillierte Beschreibung der Methode s. Abschnitt 4.5.2). Aus den Untersuchungen, in denen die Identifikation von Verlaufsgruppen fokussiert wurde, ergeben sich Hinweise auf das Vorhandensein mehrerer Subgruppen mit distinkten Symptomverläufen, die mit bestimmten Patientenmerkmalen sowie dem Therapieergebnis im Zusammenhang stehen (Cinar, Oude Voshaar, Janzing, Birkenhager, Buitelaar & van den Broek, 2010; Cuijpers, van Lier, van Straten & Donker, 2005; Keller & Hautzinger, 2007; Lutz et al., 2009; Lutz et al., 2007b; Stulz et al., 2007; Stulz et al., 2010; Uher, Muthén, Souery, Mors, Jaracz, Placentino et al., 2010). Auch hier weisen die Patienten, die sich im frühen Verlauf der Behandlung verbessern, relativ konsistent eine höhere Wahrscheinlichkeit für ein positives Therapieergebnis auf. So fanden Stulz et al. (2007) in einer Untersuchung der Behandlungsverläufe von 192 ambulanten Psychotherapiepatienten während der ersten sechs Behandlungswochen fünf distinkte Verlaufsgruppen. Zwei Gruppen wiesen eine ähnliche mittlere Ausgangsbelastung auf, unterschieden sich jedoch hinsichtlich der Kontinuität des Verlaufs. Die dritte Gruppe war zu Beginn der Behandlung stärker belastet, verbesserte sich aber innerhalb der 6 Wochen stark (*Early Improvement*). Die vierte Gruppe bildete die Gruppe mit der höchsten Ausgangsbelastung und einem unveränderten Verlauf. Die Gruppe mit der geringsten Ausgangsbelastung veränderte sich ebenfalls wenig im erhobenen Zeitraum. Die Patienten, die ein *Early Improvement* aufwiesen, erreichten die besten Behandlungsergebnisse (Effektstärke $d = 2.23$) bei relativ kurzen Behandlungszeiten. Die hochbelasteten Patienten, die keine Veränderungen aufwiesen, erreichten hingegen eine mittlere Effektstärke von $d = 0.30$, trotz relativ langer Behandlungszeiten. Die geringsten Effektstärken waren bei der Gruppe mit der niedrigsten Ausgangsbelastung zu finden ($d = 0.24$). Allerdings lag diese Patientengruppe schon zu Behandlungsbeginn nur marginal oberhalb des Cut-Off-Wertes des eingesetzten Instrumentes (CORE-SF; Kurzversion des Clinical Outcomes in Routine Evaluation (CORE) measures; Cahill, Barkham, Stiles, Twigg, Rees, Hardy et al., 2006). In einer Untersuchung von 162 depressiven Patienten kommen Lutz et al. (2009) zu ähnlichen Ergebnissen: Über den Zeitraum der ersten 8 Behandlungswochen ließen sich drei Verlaufsgruppen identifizieren. Die erste Gruppe zeichnete sich durch eine mittlere Ausgangsbelastung und eine moderate Verbesserung aus. Die zweite Gruppe umfasste den größten Anteil der Stichprobe (61%) und zeigte bei hoher Ausgangsbelastung eine rasche Verbesserung der Symptomatik (*Rapid Early Improvement*), gemessen mit dem BDI (Hautzinger et al., 1994). Mit einer hohen Ausgangsbelastung und einer moderaten Verbesserung wird die dritte identifizierte Gruppe beschrieben (20% der Stichprobe). Hinsichtlich des Outcomes zeigte sich, dass alle Patienten der *Rapid Early Improvement* Gruppe zum Ende der 16-wöchigen Behandlung eine reliable Verbesserung aufwiesen, die

auch über die Katamnese-Zeiträume von 6, 12 und 18 Monaten nach Behandlungsende weitgehend aufrecht erhalten werden konnten. Interessant ist auch das Ergebnis der initial gering belasteten Patienten: Hier erreichten nur 34% der Patienten eine reliable Verbesserung zum Behandlungsende, die Raten erhöhten sich in den Katamnese-Erhebungen innerhalb der Gruppe auf bis zu 65% (Lutz et al., 2009). Anhand eines umfangreichen Datensatzes von 2 206 Patienten aus einer US-amerikanischen Psychotherapieambulanz führte die gleiche Arbeitsgruppe ebenfalls GMM-Analysen zur Identifikation von Verläufen durch (Lutz et al., 2007b). Aufgrund der großen Stichprobe war es möglich, die Analysen getrennt für niedrig-, mittel- und hochbelastete Patienten durchzuführen. Für jede der drei Patientengruppen konnten vier distinkte Verlaufgruppen gefunden werden. Zusammenfassend zeigte sich für die Subgruppe der initial gering belasteten Patienten mit einer deutlichen Verbesserung im Verlauf das beste Behandlungsergebnis (Lutz et al., 2007b). In einer anderen Studien wurde der Verlauf der depressiven Symptomatik während einer 15-wöchigen elektrokonvulsiven Therapie untersucht (Cinar et al., 2010). Als Outcome-Maß wurde die HAMD-17 (Hamilton, 1960) zum Ende der Behandlung eingesetzt. Für die Stichprobe von 156 depressiven Patienten konnten fünf distinkte Symptomverlaufgruppen gefunden werden: Die drei größten Gruppen werden von den Autoren mit einem sofortigen (*fast improvement*; 25%), einem intermediärem (*intermediated improvement*, 31%) und einem langsamen Ansprechen (*slow improvement*, 20%) beschrieben. Die vierte Gruppe ist durch eine verzögerte Verbesserung (*slow improvement with delayed onset*; 12%) und die fünfte durch keine Verbesserung (*no improvement*, 13%) gekennzeichnet. In der ersten Gruppe (*fast improvement*) erfüllten fast alle Patienten das Kriterium der Response (50% Verbesserung zum Ausgangswert im HAMD) und der Remission (Unterschreiten des Cut-Off-Wertes von 7 im HAMD). In einer Untersuchung der Veränderung der depressiven Symptomatik von 807 Patienten mit einer depressiven Erkrankung, die mit unterschiedlichen Antidepressiva behandelt wurden, konnten mittels GMM zwei Verlaufgruppen identifiziert werden (Uher et al., 2010): Der Großteil der Stichprobe (75%) wies über den Erhebungszeitraum von 12 Wochen eine langsame und annähernd lineare Verbesserung der Symptomatik auf (*gradual improvement*), während die kleinere Gruppe (25%) durch den Verlauf einer raschen Verbesserung innerhalb der ersten 3 Wochen mit anschließender leichter gradueller Verbesserungen (*rapid improvement*) gekennzeichnet war. Als Outcome-Maße wurden die HAMD-17 (Hamilton, 1960) und das BDI (Beck et al., 1961; Hautzinger et al., 1994) eingesetzt. In der Gruppe *rapid improvement* wurden bedeutend höhere Responseraten (50% Verbesserung im HAMD-17 im Vergleich zum initialen Wert im HAMD-17) und Remissionsraten (Unterschreiten des Cut-Off-Wertes von 7 im HAMD-17 oder Unterschreiten des Cut-Off-

Wertes von 8 im BDI) zum Ende der Behandlung erreicht als in der *gradual improvement* Gruppe (Uher et al., 2010).

In den wenigen Studien, in denen der *gesamte Symptomverlauf während der stationären Behandlung* untersucht wurde, finden sich erste Hinweise auf ähnliche Verlaufsmuster der Symptombelastung, wie sie für die ambulante Psychotherapie gefunden wurden (de Jong-Meyer, Hautzinger, Rudolf, Strauß & Frieck, 1996; Drymalski & Washburn, 2011; Hautzinger et al., 1996b; Keller & Hautzinger, 2007; Schauenburg, Sammet & Strack, 2001). Das Ansprechen auf die Behandlung innerhalb der ersten 2 bis 3 Wochen scheint demnach auch in diesem Setting ein relevanter Prädiktor für das Outcome bei Therapieende zu sein (Drymalski & Washburn, 2011; Hautzinger et al., 1996b; Schauenburg et al., 2001). Keller und Hautzinger (2007) analysierten in einer Studie die Verläufe von 203 depressiven Patienten, die sowohl ambulante als auch stationäre Behandlungen einschloss. Mittels GMM konnten die Autoren vier Verlaufgruppen identifizieren: Die größte Gruppe, die 65% der Patienten umfasst, war durch eine mittlere Ausgangsdepressivität und eine langsame, aber stetige Besserung gekennzeichnet. Die zweite Gruppe wies eine hohe initiale Belastung und eine geringe Veränderungen auf (26%). Die beiden verbleibenden Gruppen umfassten relativ wenige Patienten und können durch einen hohen Ausgangswert und eine verzögerte, zunehmend schnellere Besserung (6%) und durch eine ähnlich hohe Ausgangsbelastung, aber einer rapiden Besserung in den ersten 3 Wochen beschrieben werden (3%). Eine Remission am Ende der Behandlung, operationalisiert anhand des Unterschreitens des Cut-Off-Wertes im BDI (Hautzinger et al., 1994), zeigten mehr Patienten der letzten beiden Gruppen als Patienten der erstgenannten Gruppen. Die Befunde von Drymalski und Washburn (2011), die erstmals die Symptomverläufe depressiver Patienten während der teilstationären Behandlung untersuchten, stützen die Annahme einer Verlaufgruppe mit einer Early Response. Zwar wurden hier a priori zwei Gruppen hinsichtlich des Auftretens bzw. Ausbleibens von Sudden Gains definiert, die mittleren Verläufe beider Gruppen lassen dennoch auf eine Gruppe mit mit einem graduellen Verlauf und ein zweite Gruppe mit einer höheren initialen Symptombelastung und dem Verlauf einer Early Response schließen.

2.6 Vorhersage der Behandlungsdauer und des Behandlungsverlaufs

Im Folgenden wird der aktuelle Forschungsstand zu den Prädiktoren der Behandlungsdauer und des Symptomverlaufs dargestellt. Die empirischen Befunde sollen dazu beitragen, Annahmen zur Vorhersage des Symptomverlaufs abzuleiten. Der erste Abschnitt beinhaltet Ergebnisse zur Vorhersage der Behandlungsdauer, die ergänzend zu den im zweiten Abschnitt beschriebenen Prädiktoren des Symptomverlaufs herangezogen werden.

2.6.1 Vorhersage der Behandlungsdauer

Im Zuge der Untersuchung der optimalen Therapiedauer wurde deutlich, dass die Dauer einer Psychotherapie von einer Vielzahl an Patienten- und Behandlungsmerkmalen abhängig ist (für einen Überblick vgl. Lambert & Ogles, 2004). Häufig stehen dabei Patientenmerkmale, insbesondere krankheitsbezogene und soziodemographische Variablen, im Fokus. Krankheitsbezogene Variablen, die in einem Zusammenhang mit der Behandlungsdauer stehen, sind vor allem die *Diagnose* und Faktoren, die weitgehend die *Krankheitsschwere* und *Chronifizierung* abbilden (vgl. Anderson & Lambert, 2001; Bassler, Krauthauser & Hoffmann, 1995a; Borgart & Meermann, 1999; Neeb et al., 2001; Zielke, 2006; Zielke, Dehmlow, Wülbeck & Limbacher, 1997). Je schwerer und langanhaltender die Erkrankung ist, operationalisiert durch eine Vielzahl an Variablen wie Ausgangsbelastung (Anderson & Lambert, 2001), psychische Komorbidität (Neeb et al., 2001; Zielke, 2006), Krankheitsdauer (Zielke, 2006; Zielke et al., 1997), Anzahl der Arbeitsunfähigkeitstage, Anzahl der behandlungsrelevanten Problembereiche, Schweregradbeurteilung aus Patientensicht (Zielke et al., 1997) sowie vorangehende stationäre oder ambulante psychotherapeutische Behandlung (Bassler et al., 1995a; Zielke et al., 1997), desto höher ist die erforderliche Mindestdosis an Therapiesitzungen und desto länger ist der Zeitraum, bis klinisch relevante Verbesserungen eintreten. Von dem Diagnosespektrum, das in der psychosomatisch/psychotherapeutische Krankenhausversorgung sowie in der medizinische Rehabilitation des Indikationsbereiches psychische/psychosomatische Erkrankungen vorliegt, finden sich für die Diagnosegruppen depressiver Erkrankungen und Anpassungsstörungen durchschnittlich kürzere Behandlungszeiten. Die längsten Verweildauern sind bei Zwangs-, Ess- und Persönlichkeitsstörungen zu beobachten (Bassler et al., 1995a; von Heymann, Zaudig & Tritt, 2003; Zielke, 2006; Zielke et al., 1997). Steffanowski et al. (2007) untersuchen in ihrer Metaanalyse die Vorhersagbarkeit der Therapiedauer und finden als stärkste Prädiktoren das Publikationsjahr der Studie sowie die Dauer der Erkrankung. Ersteres spiegelt die allgemeine Verkürzung der stationären Behandlungsdauer in den letzten Jahren wider. In einer von der Autorin kürzlich durchgeführten systematischen Literaturanalyse zu Prädiktoren der stationären Behandlungsdauer wurden 21 Studien mit 28 untersuchten soziodemographischen und klinischen Prädiktoren analysiert (Melchior, Hergert, Hofreuter-Gätgens, Bergelt, Morfeld, Schulz et al., 2010). Die Ergebnisse der Primärstudien wurden unter Berücksichtigung der methodischen Qualität der Studien systematisch aufbereitet. Es zeigte sich, dass sich längere Behandlungszeiten bei allein lebenden Patienten sowie bei höheren Ausprägungen der Variablen Ausgangsbelastung, Krankheitsdauer, Anzahl der Vorbehandlungen, Arbeitsunfähigkeit und psychische Komorbidität finden. Im Bezug auf die Diagnose zeigte sich, dass Patienten mit Zwangs-, Ess- und Persönlichkeitsstörungen länger behandelt

werden als Patienten mit depressiven Störungen, Angst-, somatoformen und Anpassungsstörungen. Des Weiteren geht eine hohe Therapiemotivation mit einer längeren Behandlung einher. Uneinheitlich sind die Befunde zu den untersuchten Prädiktoren Geschlecht, Alter, Schulbildung und Erwerbsstatus.

Weit seltener als Patientenmerkmale wird der Einfluss von Behandlungsmerkmalen auf die Behandlungsdauer in der stationären psychotherapeutischen Behandlung untersucht. Bassler (1995a) konnte einen Zusammenhang zwischen einer positiven therapeutischen Beziehung, patientenseitig eingeschätzt, und der Behandlungsdauer aufzeigen.

Häufig wird in Studien, in denen verschiedene Prädiktoren der Behandlungsdauer untersucht werden, von einer für den Therapieerfolg angemessenen Dauer ausgegangen, wobei das Therapieergebnis als kritische Größe nicht erhoben wird. Für die Rehabilitation psychischer/psychosomatischer Erkrankungen gibt es einige wenige Studien, die den Zusammenhang von Behandlungsdauer und -ergebnis unter Berücksichtigung verschiedener Patienten- und Behandlungsmerkmale betrachten (Nosper, 1999a; Schmitz-Buhl et al., 1999). Die Güte des Prozesserlebens, patientenseitig eingeschätzt, erwies sich in der Untersuchung von Nosper (1999a) ab der dritten Behandlungswoche als ausschlaggebend für den Therapieerfolg. Ein konstruktives Prozesserleben begünstigt demnach den Dauer-Wirkungs-Zusammenhang nach einer gewissen Behandlungsdauer positiv. Schmitz-Buhl et al. (1999) zeigen, dass die Vorhersage des Behandlungserfolgs durch die Behandlungsdauer von der Ausprägung der Variablen Beschwerdenstärke und Therapiemotivation abhängig ist.

2.6.2 Vorhersage des Symptomverlaufs während der Behandlung

Aussagen zum Zusammenhang von Patienten- bzw. Behandlungsmerkmalen mit bestimmten *Verlaufsmustern der Symptombelastung* sind aufgrund der geringen Studienlage schwer zu treffen. In der beschriebenen Studie von Lutz et al. (2009) wurden die Symptomverläufe von 162 depressiven Patienten untersucht. Ein frühes Ansprechen geht mit reliablen Verbesserungen bei Therapieende sowie 6, 12 und 18 Monaten nach Behandlung einher. Zusätzlich wurden die Art der Behandlung (kognitive Verhaltenstherapie vs. interpersonelle Psychotherapie vs. Pharmakotherapie) und die allgemeine symptomatische Ausgangsbelastung als Prädiktoren des Behandlungsergebnisses untersucht. Die Ergebnisse zeigen, dass das Verlaufsmuster – neben der Ausgangsbelastung – einen höheren prädiktiven Wert für das Behandlungsergebnis zu haben scheint als die Art der erhaltenen Behandlung. Als weitere Prädiktoren der Verlaufsmuster bzw. des Behandlungsergebnisses wären motivationale Faktoren und weitere Merkmale der Erkrankungsschwere denkbar. In der oben beschriebenen Untersuchung von Stulz et al. (2007) erweisen sich ein höheres Alter und eine erhöhte

Ängstlichkeit zu Behandlungsbeginn als signifikante Prädiktoren einer frühen Verbesserung. Für Geschlecht, Depressivität und interpersonelle Probleme konnte kein Zusammenhang zum (frühen) Behandlungsverlauf gefunden werden (Stulz et al., 2007). In der beschriebenen Analyse von 2 206 ambulanten Psychotherapiepatienten stellen die klinischen Ausgangsfaktoren depressive und interpersonelle Probleme, agoraphobische Symptome und Somatisierung sowie Probleme im Arbeitsumfeld signifikante Prädiktoren der Zugehörigkeit zu den 12 identifizierten Verlaufgruppen dar (Lutz et al., 2007b). In den von Stulz et al. (2010) durchgeführten Reanalysen der ambulanten Behandlung chronisch depressiver Patienten zeigte sich aus einer Vielzahl untersuchter Prädiktoren, dass neben der initialen depressive Symptomatik auch die Angstsymptomatik, die vorangegangene Behandlung mit Antidepressiva und die familiären Funktionsstrukturen eine Vorhersage zu den drei identifizierten Symptomverlaufgruppen erlaubten (Anteil erklärter Varianz: 61%). In der Untersuchung der Symptomverläufe von depressiven Patienten, die über 15 Wochen mittels elektrokonvulsiver Therapie behandelt wurden, ließen sich überzufällig viele Patienten mit einer Persönlichkeitsstörung als Ko-Diagnose in der Verlaufgruppe *no improvement* finden. Gleichzeitig wiesen die wenigsten Patienten der Gruppe *fast improvement* eine Therapieresistenz einer pharmakotherapeutischen Behandlung auf. Die Prädiktoranalysen erfolgten hier allerdings nur univariat, Interaktionseffekte der Prädiktoren wurden nicht berücksichtigt (Cinar et al., 2010). In der Studie von Uher et al. (2010) wurde der Verlauf der Symptombelastung von depressiven Patienten unter antidepressiver Behandlung untersucht. Die mittels GMM identifizierten Gruppen *rapid improvement* (25% der Stichprobe) und *gradual improvement* (75%) unterschieden sich im Alter und in der Dauer der gegenwärtigen depressiven Episode: Patienten, die innerhalb der ersten 3 Wochen eine deutliche Verbesserung der Symptomatik zeigten, waren signifikant jünger und wiesen eine kürzere Dauer der depressiven Episode auf als die Patienten mit einem langsamen, stetigen Verbesserungsverlauf. Keine Unterschiede zwischen den beiden Gruppen ließen sich in Hinblick auf das Geschlecht, den Familienstand, die Anzahl der Kinder, den Erwerbsstatus, die initiale depressive Symptomatik und die Anzahl vorangegangener depressiver Episoden finden. Auch hier wurden die Merkmale ausschließlich im Einzelnen analysiert, das potentielle Zusammenwirken von Prädiktormerkmalen im Bezug auf den Zusammenhang zum Verlauf wurde nicht untersucht.

Zusammenfassend zeigt sich, dass bislang wenige Befunde zu a) *der Identifikation differenzierter Symptomverläufe während der stationären Behandlung*, b) *den Prädiktoren der Symptomverläufe* und c) *dem Zusammenhang zwischen Symptomverläufen und dem längerfristigen Behandlungsergebnis* vorliegen. Begründet ist dies unter anderem durch die in der klinischen Routine der stationären Behandlung schwierig zu realisierenden

Verlaufsmessungen. Die Ergebnisse zum Behandlungsprozess aus der ambulanten Psychotherapieforschung sind nur sehr bedingt auf das stationäre psychotherapeutische Setting übertragbar, da sich sowohl die Behandlungsinhalte und -modalitäten, als auch die Dosis und Dichte sowie potentielle unspezifische Wirkfaktoren zwischen den beiden Settings unterscheiden. Es wird deutlich, dass ein hoher Bedarf an systematischem empirischem Wissen zum Zusammenhang von Verlauf und Ergebnis der stationären psychotherapeutischen Behandlung besteht. Untersuchungen zu dieser Fragestellung sollten im Hinblick auf die externe Validität in multizentrischen Studien und, im Hinblick auf die Stabilität und Nachhaltigkeit der Veränderungen unter Einbezug eines längerfristigen Outcomes, d.h. unter Einbezug von Follow-up-Messungen, durchgeführt werden (Lambert & Ogles, 2004).

Auf der Grundlage des derzeitigen Forschungsstandes zum psychotherapeutischen Prozess lassen sich zusammenfassend dennoch einige Annahmen ableiten: In allen Studien, in denen der Behandlungsprozess durch die Ausschöpfung des gesamten Verlaufs dargestellt wurde, ließen sich mindestens zwei distinkte Verlaufgruppen identifizieren (Drymalski & Washburn, 2011; Keller & Hautzinger, 2007; Lutz et al., 2009; Lutz et al., 2007b; Stulz et al., 2007; Stulz et al., 2010). Hinsichtlich der Anzahl der identifizierten Gruppen sowie der Anzahl der Patienten, die den jeweiligen Gruppen zugeordnet wurden, lassen sich keine einheitlichen Ergebnisse finden. Bemerkenswert scheint jedoch, dass sich in fast allen Studien eine Subgruppe identifizieren lässt, die sich schon früh im Verlauf verbessert und anschließend dieses Niveau bis zum Behandlungsende hält (im Sinne einer negativ beschleunigten Kurve). Diese Verlaufgruppe wird von den Autoren mit den Begriffen *Early* oder *Rapid Improvement* bzw. *Early Response* benannt. Diese Gruppe erzielte konsistent das beste Behandlungsergebnis, wobei das Outcome nur selten auch längerfristig untersucht wurde. Zur Vorhersage spezifischer Verlaufsmuster durch Patienten- und Prozessmerkmale lassen sich aufgrund der geringen Befundlage keine Annahmen ableiten. Infolgedessen kann jedoch auf die Ergebnisse zur Prädiktion der Behandlungsdauer und des Behandlungsergebnisses zurückgegriffen werden, um Annahmen zu potentiellen Prädiktoren des Symptomverlaufs zu generieren. Einschränkend ist darauf hinzuweisen, dass der größte Teil der Untersuchungen zur Early Response oder Symptomverlaufgruppen in der Psychotherapie auf die ambulante Behandlung von depressiv erkrankten Patienten, meist mittels kognitiver Verhaltenstherapie bezogen ist (Drymalski & Washburn, 2011; Keller & Hautzinger, 2007; Lambert, 2005; Lutz et al., 2009; Stulz et al., 2010; Tang & DeRubeis, 1999a; Tang & DeRubeis, 1999b; Tang, Luborsky & Andrusyna, 2002; Uher et al., 2010).

Schlussfolgernd wird ein hoher Bedarf an verbesserten methodischen Ansätzen zur Analyse von Zeitverläufen, die Patientensubgruppen mit unterschiedlichen Symptomverläufen darstellen, ebenso wie der Einbezug der längerfristigen Effekte und der Analyse der Vorhersage des Verlaufs, deutlich (Lambert, 2005; Wilson, 1999).

3 Fragestellungen und Hypothesen

3.1 Zielsetzung der Untersuchung

In der stationären Versorgung psychisch erkrankter Patienten fehlt die empirische Basis zur Bestimmung angemessener Behandlungszeiten. Die Optimierung von Behandlungsverläufen und -zeiten besitzt eine hohe Relevanz im Sinne des Anrechts von Patienten auf eine effektive und effiziente Behandlung mit schnellst- und größtmöglicher Minimierung des Leidens sowie – unter gesellschaftlicher und ökonomischer Perspektive – einer gerechten Verteilung vorhandener Ressourcen (Lutz, Grawe & Tholen, 2003).

Ziel der Arbeit ist die Analyse von Symptomverläufen und deren Vorhersagekraft für das längerfristige Behandlungsergebnis unter besonderer Berücksichtigung von Dauer/Dosis-Wirkungs-Zusammenhängen. Dabei wird der Frage nachgegangen, ob sich verschiedene Muster der Veränderung der Symptombelastung identifizieren lassen und ob diese eine Vorhersage bezüglich des längerfristigen Therapieergebnisses erlauben. Zu den explorativen Verlaufsanalysen sollen zusätzlich theoriegeleitete Verlaufgruppen gebildet werden, insbesondere um das Konzept der empirisch fundierten Early Response zu prüfen. Für die klinische Implikation scheint es von besonderer Bedeutung, Patienten frühzeitig zu identifizieren, die mit einer gewissen Wahrscheinlichkeit ein bestimmtes Verlaufsmuster der Symptombelastung aufweisen werden, um Behandlungsprozesse und ggf. auch den Behandlungsaufwand an individualisierte Bedarfe anzupassen (Clarkin & Levy, 2004). Daher sollen theoretisch und empirisch fundierte Prädiktoren, also soziodemographische und klinische Patientenmerkmale, hinsichtlich ihres Zusammenhangs zu den identifizierten Symptomverläufen untersucht werden. Um die Wirksamkeit der Behandlung in Abhängigkeit vom Verlauf der Symptombelastung, der Dauer und der Dosis der Behandlung zu untersuchen, wird anschließend der Zusammenhang zwischen Symptomverlauf, Behandlungsdauer, Behandlungsdosis und längerfristigem Behandlungsergebnis analysiert. Die Ergebnisse können einen wichtigen empirischen Beitrag für die Optimierung von Behandlungsprozessen liefern: Dies könnte z.B. über die Entwicklung zeitnaher Feedback-Interventionen für die Behandler zur Prognose aufgrund des aktuellen Verlaufs erfolgen. Folglich können die Ergebnisse dazu beitragen, dass relevante klinische Implikationen abgeleitet werden und dienen dadurch der verbesserten Wissenschafts-Praxis-Integration. Auf wissenschaftlicher Ebene soll zudem ein Beitrag zur Weiterentwicklung psychotherapeutischer Prozess-Outcome-Forschung durch den Einsatz innovativer Methoden zur Verlaufsanalyse (Growth Mixture Modeling, vgl. Abschnitt 4.5.2) geleistet werden.

3.2 Fragestellungen

FRAGESTELLUNG 1

Wie verläuft die Symptombelastung bei Patienten mit psychischen Erkrankungen während der stationären Behandlung?

Mit der ersten Fragestellung soll untersucht werden, ob sich Subgruppen von Patienten identifizieren lassen, die unterschiedliche Verläufe der Symptombelastung während der Behandlung aufweisen. Dabei werden zwei Kriterien zur Gruppenbildung herangezogen:

Zum einen wird mit einem explorativen Ansatz untersucht, ob und welche Patientengruppen sich aufgrund der Ähnlichkeit der Verläufe, d.h. der Ausgangsbelastung und dem Ausmaß der Veränderung im Verlauf der gesamten Behandlung, identifizieren lassen. Dabei soll die Information über den gesamten Verlauf, also alle vorhandenen Verlaufsmessungen, zur Gruppenbildung genutzt werden. Diese Gruppen sollen eine möglichst hohe Homogenität innerhalb einer Gruppe und gleichzeitig möglichst hohe Heterogenität zwischen den Gruppen hinsichtlich der Ausgangsbelastung und der Veränderungen im Verlauf aufweisen.

Als zweites Kriterium der Subgruppenbildung kommt eine theoriebasierte Definition der Response zur Anwendung. Dabei soll theoriegeleitet überprüft werden, welcher Anteil an Patienten nach diesem Kriterium mit a priori festgelegten Response-Zeitpunkten während bzw. am Ende der Behandlung zuzuordnen ist. Anschließend werden beide Subgruppenbildungen miteinander verglichen. Es ergeben sich folgende drei Unterfragestellungen:

A Explorative Verlaufsanalyse: Lassen sich Patientensubgruppen mit ähnlichen Verläufen der Symptombelastung, definiert über die Ausgangsbelastung und das Ausmaß sowie die Richtung der Veränderung über die gesamten Behandlungszeitraum, identifizieren?

Welche Verläufe (Ausgangsbelastung, Ausmaß und Richtung der Veränderung) lassen sich auf der Basis des explorativen Ansatzes identifizieren?

Welcher Anteil an Patienten ist den gefundenen Gruppen zugeordnet?

B Theoriegeleitete Verlaufsanalyse: Wie viele Patienten werden den folgenden drei, nach theoriebasierter Definition eingeteilten, a priori festgelegten Verlaufgruppen zugeordnet?

- a) Early Response: Es tritt eine relevante Verbesserung der Symptombelastung nach der zweiten oder dritten Woche im Vergleich zur Ausgangsbelastung (Symptombelastung bei Aufnahme) auf.
- b) Delayed Response: Es tritt eine relevante Verbesserung der Symptombelastung zur Entlassung im Vergleich zur Ausgangsbelastung (Symptombelastung bei Aufnahme) auf, nicht jedoch innerhalb der ersten 3 Wochen.
- c) Non-Response: Es tritt weder innerhalb der ersten 3 Wochen, noch zur Entlassung eine relevante Verbesserung der Symptombelastung auf.

C Vergleich von explorativen und theoriegeleiteten Verlaufgruppen:

Wie verteilen sich die Patienten zwischen den identifizierten explorativen und theoriegeleiteten Verlaufgruppen? Welche Entsprechung finden die jeweiligen Gruppen untereinander?

Zusätzlich zu den Subgruppenanalysen wird für die Gesamtstichprobe in Fragestellung 1 D untersucht, zu welchem Zeitpunkt während der Behandlung Verbesserungen im primären Outcome *Symptombelastung* nach vorab definierten Kriterien auftreten. Diese deskriptiven Analysen sollen dazu beitragen, ein differenziertes Abbild des Verlaufs der Symptombelastung während der Behandlung nach gängigen in der Literatur eingesetzten Kriterien zu erhalten.

D Zu welchem Zeitpunkt erreichen wie viele Patienten eine Verbesserung in den folgenden drei Kriterien?

- 1) Verbesserung mit einer mindestens kleinen Effektstärke
- 2) Verbesserung um mindestens 50% der Ausgangsbelastung
- 3) Klinisch signifikante Verbesserung (Verbesserung um die Mindestgröße des Reliable Change Index und unterschreiten des Cut-Off-Wertes)

FRAGESTELLUNG 2

Lässt sich der Verlauf der Symptombelastung während der Behandlung durch Patientenmerkmale sowie die Güte der therapeutischen Beziehung in der frühen Behandlungsphase vorhersagen?

Mit der zweiten Fragestellung sollen die in Fragestellung 1 identifizierten Subgruppen mit unterschiedlichen Verläufen hinsichtlich ihrer Vorhersagbarkeit durch Patientenmerkmale untersucht werden. Die Frage zielt darauf ab, zu überprüfen, inwiefern theoretisch abgeleitete Patientencharakteristika sowie die Güte der therapeutischen Beziehung in der Anfangsphase der Behandlung mit der Zugehörigkeit zu den jeweiligen Verlaufsgruppen zusammenhängen. Lassen sich in Fragestellung 1 A keine homogenen Subgruppen identifizieren ist diese Fragestellung irrelevant.

Ausgehend von den Ergebnissen von Fragestellung 1 ergeben sich folgende zwei Unterfragestellungen:

- A** Lässt sich die Zugehörigkeit zu den explorativen Verlaufsgruppen durch empirisch abgeleitete Patienten- und Prozessmerkmale vorhersagen?
- B** Lässt sich die Zugehörigkeit zu den theoriegeleiteten Verlaufsgruppen durch empirisch abgeleitete Patienten- und Prozessmerkmale vorhersagen?

Fragestellung 2 A und B werden hinsichtlich der folgenden theoretisch-empirisch abgeleiteten Patienten- und Prozessmerkmale untersucht:

Die Patientenmerkmale bilden zum einen die soziodemographischen Merkmale Alter, Geschlecht, Familienstand und Partnersituation, Schulbildung und Erwerbstätigkeit. Klinische und sozialmedizinische Merkmale stellen die Diagnose, Komorbidität, stationäre und ambulante psychotherapeutische Vorbehandlung sowie Arbeitsunfähigkeitszeiten bzw. Krankheitszeiten vor Beginn der Behandlung dar. Zusätzlich werden die personenbezogenen Merkmale Psychotherapiemotivation und die allgemeine Selbstwirksamkeitserwartung einbezogen.

Als Prozessmerkmale werden die Behandlungsdauer, die Dosis psychotherapeutischer Maßnahmen, die Dosis verhaltenstherapeutischer bzw. tiefenpsychologisch fundierter Psychotherapie-Maßnahmen sowie die behandelnde Klinik untersucht. Als personenbezogenes Prozessmerkmal geht die Güte der therapeutischen Beziehung in der frühen Phase der Behandlung in die Analysen mit ein.

FRAGESTELLUNG 3**Lässt sich das längerfristige Behandlungsergebnis 6 Monate nach der Entlassung durch den Symptomverlauf während der Behandlung vorhersagen?**

Mit der dritten Fragestellung soll untersucht werden, inwiefern bestimmte Verläufe der Symptombelastung während der Behandlung mit dem längerfristigen Behandlungsergebnis zusammenhängen. Dabei wird das Behandlungsergebnis über ein primäres und ein sekundäres Ergebnismaß operationalisiert.

Lassen sich in Fragestellung 1 A keine homogenen Subgruppen identifizieren, sind die Fragestellungen 3 A und 3 C, die sich auf die explorativen Verlaufgruppen beziehen, irrelevant.

Ausgehend von den Ergebnissen aus Fragestellung 1 ergeben sich folgende vier Unterfragestellungen:

- A** Wie gut erfolgt eine Vorhersage des längerfristigen Behandlungsergebnisses, operationalisiert über das primäre Outcome *Symptombelastung* 6 Monate nach Entlassung, durch die explorativen Verlaufgruppen, unter Kontrolle von Patienten- und Prozessmerkmalen?
- B** Wie gut erfolgt eine Vorhersage des längerfristigen Behandlungsergebnisses, operationalisiert über das primäre Outcome *Symptombelastung* 6 Monate nach Entlassung, durch die theoriegeleiteten Verlaufgruppen, unter Kontrolle von Patienten- und Prozessmerkmalen?
- C** Wie gut erfolgt eine Vorhersage des längerfristigen Behandlungsergebnisses, operationalisiert über das sekundäre Outcome *psychische Aspekte der gesundheitsbezogenen Lebensqualität* 6 Monate nach Entlassung, durch die explorativen Verlaufgruppen, unter Kontrolle von Patienten- und Prozessmerkmalen?
- D** Wie gut erfolgt eine Vorhersage des längerfristigen Behandlungsergebnisses, operationalisiert über das sekundäre Outcome *psychische Aspekte der gesundheitsbezogenen Lebensqualität* 6 Monate nach Entlassung, durch die theoriegeleiteten Verlaufgruppen, unter Kontrolle von Patienten- und Prozessmerkmalen?

Fragestellung 3 A bis D werden unter Berücksichtigung empirisch begründeter Patienten- und Prozessmerkmale, die in einem potentiellen Zusammenhang mit dem Behandlungsergebnis stehen, untersucht.

Als Patientenmerkmale werden die folgenden soziodemographischen und klinischen Merkmale einbezogen, die schon in Fragestellung 2 als Prädiktoren untersucht wurden: Alter, Geschlecht, Familienstand und Partnersituation, Schulbildung, Erwerbstätigkeit, Diagnose, Komorbidität, stationäre und ambulante psychotherapeutische Vorbehandlung und Arbeitsunfähigkeitszeiten bzw. Krankheitszeiten vor Beginn der Behandlung. Zusätzlich werden die personenbezogenen Merkmale Psychotherapiemotivation und die allgemeine Selbstwirksamkeitserwartung einbezogen.

Als Prozessmerkmale werden die Behandlungsdauer, die Dosis psychotherapeutischer Maßnahmen sowie die anteilige Dosis verhaltenstherapeutischer bzw. tiefenpsychologisch fundierter Psychotherapie-Maßnahmen untersucht. Als personenbezogenes Prozessmerkmal geht die Güte der therapeutischen Beziehung in der frühen Phase der Behandlung in die Analysen mit ein. Zur Berücksichtigung potentieller Zusammenhänge, die durch die behandelnde Klinik bedingt werden, wird die Klinik ebenfalls als Prädiktor zur Vorhersage des längerfristigen Behandlungsergebnisses eingeschlossen.

FRAGESTELLUNG 4

Durch welche Verlaufsguppen (explorativ oder theoriegeleitet) erfolgt eine bessere Vorhersage des katamnestischen Behandlungsergebnisses (primäres und sekundäres Outcome)?

3.3 Hypothesen

In *Fragestellung 1 A* sollen mit einer explorativen Vorgehensweise unterschiedliche Verlaufsguppen der Symptombelastung identifiziert werden. Aufgrund des explorativen Charakters der Untersuchungsfragestellung und der geringen empirischen Basis zum Verlauf der Symptombelastung im stationären Setting wird keine Hypothese formuliert.

In der *Fragestellung 1 B* werden Symptomverlaufsguppen a priori über zeitliche und qualitative Kriterien definiert. Es wird angenommen, dass die theoretisch und empirisch am besten fundierte Gruppe mit einer Early Response (vgl. Abschnitt 2.5.5) nach den hier formulierten theoriegeleiteten Kriterien vorzufinden ist. Dabei wird angenommen, dass ein

Mindestanteil von 5% der Stichprobe das Kriterium einer Early Response erreicht. Hieraus leitet sich die erste Hypothese ab:

Hypothese 1:

Für die nach theoriegeleiteten Kriterien definierte Verlaufsgruppe „Early Response“ lässt sich ein Anteil von mindestens 5% der Stichprobe finden.

Für *Fragestellung 1 C*, in der die explorativen und die theoriegeleiteten Verlaufsgruppen miteinander verglichen werden, wird aufgrund des explorativen Charakters von *Fragestellung 1 A* keine Hypothese formuliert. Es wird jedoch angenommen, dass es eine Übereinstimmung zwischen den theoriegeleiteten und den explorativen Gruppen gibt.

In der *Fragestellung 1 D* wird die jeweilige Anzahl an Patienten, die sich im wöchentlichen Verlauf anhand von drei Kriterien, die das Ausmaß einer Veränderung im Verlauf festlegen, dargestellt. Hier kann ebenfalls aufgrund der geringen empirischen Basis zu wöchentlichen Verläufen im stationären Setting keine Hypothese aufgestellt werden.

Mit der *zweiten Fragestellung* (*Fragestellung 2 A* und *2 B*) soll untersucht werden, ob sich die Gruppenzugehörigkeit zu den in *Fragestellung 1* gebildeten Verlaufsgruppen durch empirisch abgeleitete Patienten- und Prozessmerkmale vorhersagen lässt. Als potentielle Prädiktoren der Verlaufsgruppenzugehörigkeit werden Merkmale ausgewählt, für die Hinweise auf Zusammenhänge mit dem Symptomverlauf (vgl. Abschnitt 2.6.2), mit der Behandlungsdauer (vgl. Abschnitt 2.6.1) oder mit dem Behandlungsergebnis (vgl. Abschnitt 2.4.1) vorliegen. Obwohl es sich bei den empirischen Befunden somit auch um Prädiktoren des Outcomes und der Behandlungsdauer und nicht ausschließlich um Prädiktoren des Verlaufs handelt, wird angenommen, dass diese Merkmale potentiell auch in einem Zusammenhang mit der Symptombelastung im Verlauf, und demnach mit den Verlaufsgruppen, stehen. Zur Vorhersage der Gruppenzugehörigkeit werden für die explorativen und die theoriegeleiteten Verlaufsgruppen folgende Hypothesen abgeleitet:

Hypothese 2:

Durch das vorgegebene Set an potentiellen Prädiktoren (Patienten- und Prozessmerkmale) wird ein signifikanter Anteil der Varianz in den explorativen Verlaufsgruppen erklärt.

Hypothese 3:

Durch das vorgegebene Set an potentiellen Prädiktoren (Patienten- und Prozessmerkmale) wird ein signifikanter Anteil der Varianz in den theoriegeleiteten Verlaufgruppen erklärt.

Hinsichtlich *Hypothese 2* und *3* wird weiterhin angenommen, dass die Höhe der Varianzaufklärung mindestens einer mittleren Effektstärke entspricht. Für die Beurteilung der Güte des Gesamtmodells wird entsprechend ein Mindestwert von 40% erklärter Varianz herangezogen (vgl. Tabelle 3, Abschnitt 4.5.4, Seite 83). Die Annahmen zur Höhe der Effektstärken werden jedoch nicht als Hypothesen formuliert, da die empirische Fundierung zur genauen Einschätzung der erklärten Varianz in Symptomverlaufgruppen zu gering ist.

Die *dritte Fragestellung* untersucht den Vorhersagewert der identifizierten Verlaufgruppen für das katamnestische Behandlungsergebnis hinsichtlich des primären Outcomes *Symptombelastung* und des sekundären Outcomes *psychische Aspekte der gesundheitsbezogenen Lebensqualität* zum Zeitpunkt der 6-Monats-Katamnese. Es wird angenommen, dass zwischen den identifizierten Verlaufgruppen, unter Kontrolle des Ausgangswertes des jeweiligen Outcome-Parameters, bedeutsame Unterschiede im primären und sekundären Outcome bestehen. Zudem besteht die Annahme, dass unter Berücksichtigung und statistischer Kontrolle empirisch basierter Prädiktoren des Behandlungsergebnisses (Patienten- und Prozessmerkmale) noch ein bedeutender zusätzlicher Anteil der Gesamtvarianz im jeweiligen Outcome-Parameter durch die Verlaufgruppen erklärt werden kann. Folglich werden jeweils zwei Hypothesen zu den explorativen Verlaufgruppen und zwei Hypothesen zu den theoriegeleiteten Verlaufgruppen formuliert. Die vier Hypothesen werden erst für das primäre Outcome *Symptombelastung zur Katamnese (Hypothese 4 bis 7)* und anschließend in der gleichen Weise für das sekundäre Outcome *psychische Aspekte der gesundheitsbezogenen Lebensqualität zur Katamnese (Hypothesen 8 bis 11)* überprüft:

Hypothese 4 (und 8):

Die explorativen Verlaufgruppen unterscheiden sich, bei Kontrolle des Ausgangswertes des Ergebnisparameters, signifikant im katamnestischen Behandlungsergebnis.

Hypothese 5 (und 9):

Durch die explorativen Verlaufsguppen kann, zusätzlich zu empirisch fundierten Prädiktoren des Outcomes, ein signifikanter inkrementeller Anteil der Gesamtvarianz im katamnestischen Behandlungsergebnis aufgeklärt werden.

Hypothese 6 (und 10):

Die theoriegeleiteten Verlaufsguppen unterscheiden sich, bei Kontrolle des Ausgangswertes des Ergebnisparameters, signifikant im katamnestischen Behandlungsergebnis.

Hypothese 7 (und 11):

Durch die theoriegeleiteten Verlaufsguppen kann, zusätzlich zu empirisch fundierten Prädiktoren des Outcomes, ein signifikanter inkrementeller Anteil der Gesamtvarianz im katamnestischen Behandlungsergebnis aufgeklärt werden.

Aufgrund der geringen empirischen Basis werden Aussagen zur Höhe der Effektstärken für die Unterschiedshypothesen und die Hypothesen zum Zuwachs an Varianzaufklärung nicht in die Hypothesenformulierung aufgenommen. Dennoch wird angenommen, dass die Effektstärken der Unterschiede zwischen den jeweiligen Verlaufsguppen im jeweiligen Outcome die Höhe einer mindestens mittleren Effektstärke erreichen. Ferner wird angenommen, dass der inkrementelle Zuwachs an Varianzaufklärung durch die jeweiligen Verlaufsguppen im jeweiligen Outcome mindestens eine kleine bis mittlere Effektstärke erreicht.

In der *vierten Fragestellung* werden die explorativen und die theoriegeleiteten Verlaufsguppen hinsichtlich ihres inkrementellen Vorhersagewertes auf das katamnestische Outcome miteinander verglichen. Ausgehend von Fragestellung 3 bezieht sich der Vergleich sowohl auf das primäre als auch auf das sekundäre Outcome. Aufgrund der fehlenden empirischen Basis wird diese Fragestellung explorativ untersucht. Es wird keine Hypothese formuliert.

4 Methode

4.1 Studiendesign

In einer multizentrischen Erhebung in drei Fachkliniken zur Behandlung psychisch erkrankter Patienten (Medizinische Rehabilitation und Krankenhausbehandlung) wurden Verläufe der Symptombelastung sowie das Behandlungsergebnis der stationären psychotherapeutischen Behandlung¹³ untersucht. Dazu wurden das primäre und sekundäre Outcome zur Aufnahme (Prä; t0), zur Entlassung (Post; t1) sowie 6 Monate nach Entlassung (Follow-up; t2) mittels patientenseitiger Selbsteinschätzung erhoben. Zusätzlich zu den drei Erhebungszeitpunkten wurden während der gesamten Behandlung wöchentliche Verlaufsmessungen der patientenseitigen Symptombelastung erhoben (vgl. Abbildung 1). Die Prozesserhebung war an eine, in den Kliniken regelmäßig (wöchentlich) stattfindende Veranstaltung/Maßnahme gekoppelt, sodass ein wöchentliches Zeitintervall für die gesamte Aufenthaltsdauer in den Kliniken für den Großteil der Patienten gewährleistet werden konnte.

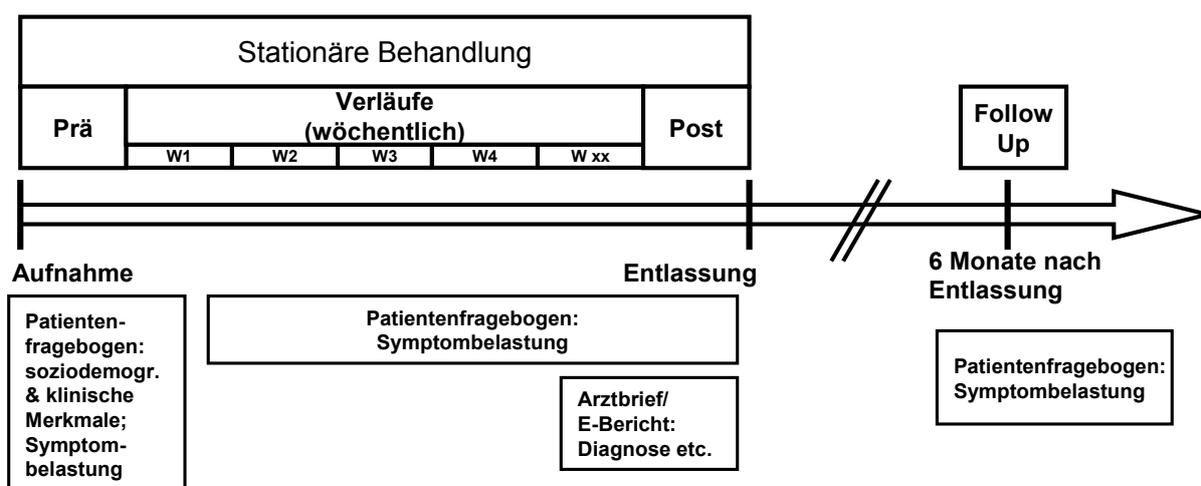


Abbildung 1: Studiendesign

¹³ Die Benennung der Behandlung als *psychotherapeutisch* ist nicht ganz präzise, da die medizinische Rehabilitation und die Krankenhausbehandlung psychischer und psychosomatischer Erkrankungen durch ein multimodales Behandlungskonzept gekennzeichnet ist (vgl. Abschnitt 2.1). Im Fokus stehen dabei zwar die psychotherapeutischen Maßnahmen, welchen Stellenwert die Psychotherapie bei der Erklärung der Wirksamkeit der stationären Behandlung einnimmt, bleibt unklar. Im Folgenden wird dennoch der Begriff der *psychotherapeutischen Behandlung* verwendet, da zum einen davon ausgegangen wird, dass psychotherapeutische Behandlungen einen wesentlichen Beitrag zum Behandlungsergebnis leisten. Zum Anderen wird der Begriff aus Gründen der besseren Lesbarkeit verwendet. Eine Annäherung an die Frage nach dem Stellenwert der Psychotherapie findet in der vorliegenden Arbeit durch die Untersuchung der Dosis psychotherapienaher Maßnahmen statt (vgl. Abschnitt 4.3.3.3).

Zusätzlich zu den Verlaufsdaten wurde ein breites Spektrum soziodemographischer und klinischer Merkmale (z.B. Diagnose) erhoben, die unter anderem aus Arztbriefen und Entlassberichten entnommen wurden (vgl. Abschnitt 4.3.3).

4.2 Durchführung der Datenerhebung

4.2.1 Setting

Die Datenerhebung erfolgte in drei Fachkliniken für Psychosomatische Medizin und Rehabilitation psychischer/psychosomatischer Erkrankungen (MediClin Seepark Klinik Bad Bodenteich, Klinik für psychosomatische Medizin der Segeberger Kliniken, Curtius Klinik Bad Malente). Bei der Auswahl der Kliniken wurde ein repräsentatives Abbild der Versorgungssituation psychisch erkrankter Patienten über die Klinikmerkmale Setting, psychotherapeutischer Schwerpunkt hinsichtlich der Therapieschulen und Bettenzahl angestrebt. Eine der drei Kliniken zeichnete sich durch ein gemischtes Spektrum an Reha- und Akutbehandlungen aus. In der zweiten Klinik wurden ausschließlich Patienten aus der Abteilung für psychosomatische Rehabilitationsmedizin rekrutiert. In der dritten Klinik war der Behandlungsschwerpunkt (und die Bettenzahl) mehrheitlich für Akutpatienten ausgerichtet. Generell entsprechen die Therapiekonzepte in allen drei Kliniken weitgehend dem multimodalen Konzept der Rehabilitationsbehandlung mit dem Fokus auf den psychotherapeutischen Behandlungen. Durch die Wahl dieser drei Fachkliniken sollte zum einen die externe Validität erhöht, zum anderen auch die Vergleichbarkeit hinsichtlich des grundsätzlichen Behandlungskonzepts gewährleistet werden. Hinsichtlich der psychotherapeutischen Ausrichtung wurden die Kliniken ebenfalls nach Kriterien der Repräsentativität ausgewählt. Die erste der drei Kliniken legt einen Schwerpunkt auf verhaltenstherapeutische Therapieverfahren, die zweite zeichnet sich durch einen integrativen-schulenübergreifenden Behandlungsansatz aus und in der dritten Klinik wird verstärkt nach tiefenpsychologisch fundierten Therapieverfahren behandelt. Die psychotherapeutischen Behandlungsmodalitäten zeichnen sich neben Einzeltherapien durch einen essentiellen Anteil an Gruppentherapien aus und spiegeln dahingehend die stationäre psychotherapeutische Routineversorgung dieses Systems wider (Barghaan, Schulz, Koch & Watzke, 2009; Watzke et al., 2009).

Das in den Kliniken vorzufindende Diagnosespektrum umfasst hauptsächlich die ICD-10-Diagnosegruppen F3 bis F6 und ist somit weitgehend repräsentativ für die beiden Versorgungsbereiche Rehabilitation psychischer/psychosomatischer Erkrankungen und psychosomatische/psychotherapeutische Krankenhausbehandlung (vgl. Abschnitt 2.1.1). Allerdings besteht in zwei der drei Kliniken ein Behandlungsschwerpunkt in den

Indikationsbereichen der Essstörungen, der für die untersuchten Behandlungssettings weniger repräsentativ ist. Dieser Aspekt muss bei der Interpretation der Ergebnisse berücksichtigt werden.

Die drei Kliniken weisen Bettenzahlen von 150 bis 200 Betten auf, die durchschnittliche Anzahl liegt bei 172 Betten.

4.2.2 Datenerhebung und Datenschutz

Für die Rekrutierung und Erhebung in den Kliniken sollte gewährleistet werden, dass möglichst alle Patienten über die Studie informiert wurden. Der Aufnahme-Fragebogen sollte möglichst zeitnah an der Aufnahme, maximal aber in den ersten 3 Tagen nach der Aufnahme in der Klinik, ausgefüllt werden. Die Verlaufs-Fragebögen sollten in einem festen zeitlichen Abstand von einer Woche und der Entlass-Fragebogen möglichst zeitnah an der Entlassung, maximal 3 bis 4 Tage vor der Entlassung, von den Patienten ausgefüllt werden.

Die Information über die Studie erfolgte in den Kliniken über die aufnehmenden Ärzte und Therapeuten im Rahmen einer Begrüßungsveranstaltung, an der alle in der laufenden Woche neu aufgenommenen Patienten teilnahmen, oder über das Personal aus dem Patientenmanagement im direkten Zweierkontakt. Äußerten die Patienten ihr Interesse an einer Teilnahme an der Studie, konnten sie an einem dafür vorgesehenen Termin (zum Teil direkt nach der Begrüßungsveranstaltung) genauere Informationen über die Studie erhalten und bei Einwilligung zur Studienteilnahme den Aufnahme-Fragebogen direkt bei diesem Termin ausfüllen. Die Implementierung wöchentlicher Erhebungen stellte eine große organisatorische Herausforderung dar. Es wurden zwei mögliche Vorgehensweisen abgewogen. Die organisatorisch weniger aufwendige Variante wäre die gewesen, dem Patienten in einem regelmäßigen Zeitabstand, z.B. gekoppelt an eine wöchentlich stattfindende Behandlungsmaßnahme, den Fragebogen auszuhändigen, und diesen an einer zentralen Stelle in der Klinik wieder abgeben zu lassen. Ein wesentlicher Nachteil dieses Vorgehens ist die Unsicherheit über den genauen Zeitpunkt des Ausfüllens. Da die Verlaufsmessung hier auf einem Zeitintervall und nicht auf einem Dosisintervall beruht, ist die möglichst exakte zeitliche Einhaltung des Intervalls bedeutend. Für das kurze Zeitintervall von einer Woche schien die Gefahr einer Verzerrung der Angaben zu hoch, selbst wenn sie nur ein oder zwei Tage verspätet abgegeben würden. Da das Behandlungskonzept eine intensive und zeitlich dichte psychotherapeutische Behandlung vorsieht, könnten Effekte nicht mehr klar auf die in einem bestimmten Zeitraum erhaltenen Maßnahmen zurückgeführt werden. Daher wurde eine andere Form der wöchentlichen Datenerhebung entwickelt, bei der die Einhaltung des zeitlichen Intervalls maximal gewährleistet sein sollte: Geplant wurde ein wöchentlich stattfindender Studien-Termin, an dem alle teilnehmenden Patienten die Fragebögen unter Betreuung eines Studienleiters bzw. einer Studienleiterin ausfüllten und

direkt im Anschluss an diese abgaben. Dazu war es zum einen notwendig, einen Termin in der Woche zu finden, an dem keine anderen Behandlungsmaßnahmen oder Veranstaltungen für die Patienten vorgesehen waren, sodass alle Patienten die Möglichkeit hatten, an der Studie teilzunehmen, ohne dass die klinische Behandlungsroutine gestört wurde. Um eine möglichst regelmäßige Teilnahme zu gewährleisten, konnte der Studien-Termin, unter der Voraussetzung der Einverständniserklärung zur Studienteilnahme durch den Patienten, in den Therapieplan mit aufgenommen werden. Eine weitere Voraussetzung war die notwendige räumliche Ausstattung in den Kliniken. Durch den konsekutiven Einschluss aller Patienten musste bei einer Bettenzahl von bis zu 200 Betten und einer geschätzten Teilnahmequote von 50%-60% mit einer Teilnehmerzahl von bis zu 120 Patienten zu einem Erhebungszeitpunkt gerechnet werden. Da dafür die räumlichen und organisatorischen Voraussetzungen nicht in allen drei Kliniken gegeben waren, wurde der Studien-Termin zum Teil auf zwei Termine in der Woche gelegt, zum Teil wurde der Termin ausreichend lang angesetzt, sodass die Patienten die Möglichkeit hatten, den Termin zeitlich etwas versetzt wahrzunehmen. Weiterhin waren auch personelle Ressourcen für die Durchführung sowie die Vor- und Nachbereitung des Studien-Termins (Erfassung und Kodierung der Fragebögen, Zuordnung der kodierten Fragebögen, Bereitstellung der Materialien etc.) für den Erhebungszeitraum erforderlich. In zwei der drei Kliniken konnte jeweils Klinikpersonal aus der Pflege und dem Patientenmanagement für diese Aufgabe gewonnen werden. In der dritten Klinik wurde die Rekrutierung und Erhebung von der Autorin selbst, mit Unterstützung einer studentischen Hilfskraft, übernommen.

Der Entlass-Fragebogen wurde ebenfalls beim Studien-Termin ausgehändigt und von den Patienten vor Ort ausgefüllt. Dabei wurde darauf geachtet, dass bei möglichen Änderungen der bewilligten Behandlungsdauer die entsprechenden Patienten rechtzeitig identifiziert wurden und die Erhebungsprozedur angepasst wurde.

Der Fragebogen für die Halbjahreskatamnese wurde auf postalischem Weg direkt aus den Kliniken versendet, da die Zuordnungen der Kodierung zu den personenbezogenen Daten aus datenschutzrechtlichen Gründen nur in den Kliniken eingesehen und bearbeitet werden konnten. Wenn 4 Wochen nach dem Versand kein Rücklauf registriert werden konnte, wurde ein Erinnerungsschreiben an den Patienten gesandt.

Die Patienteninformation, die Einverständniserklärung, des Anschreiben zur Katamnese-Erhebung sowie das Erinnerungsschreiben sind im Anhang E zu finden.

Um die verschiedenen Fragebögen zuordnen zu können, wurden sie mit einer einheitlichen fortlaufenden Code-Nummer versehen. Die Bearbeitung aller Patientendaten erfolgte pseudonymisiert, das heißt ohne die Nennung von Angaben, die Rückschlüsse auf die Person zulassen. Eine Zuordnungsliste von Patientendaten und Code-Nummern sowie die unterschriebene Einwilligung wurden in der Klinik verwahrt. Somit wurde sichergestellt, dass

die Bestimmungen des Bundesdatenschutzgesetzes erfüllt sind. Die Studie wurde sowohl von der Ethikkommission der Ärztekammer Hamburg als auch von der Datenschutzabteilung der Deutschen Rentenversicherung geprüft und mit einem positiven Votum versehen.

4.2.3 Ein- und Ausschlusskriterien

Eingeschlossen wurden alle Patienten, die im Zeitrahmen einer regulären Behandlungsdauer behandelt wurden. Ziel der Untersuchung ist es, Aussagen über die Standardbehandlung im stationären Fachsetting zu treffen. Patienten, die länger als 9 Wochen oder kürzer als 3 Wochen behandelt werden, nehmen zusammengefasst weniger als 5% der Patienten der stationären Rehabilitationsbehandlung ein (vgl. Abschnitt 2.1.2). Um Verzerrungen der Ergebnisse aufgrund spezifischer Patientengruppen oder Behandlungsmodalitäten, die eine besonders kurze oder besonders lange Behandlungsdauer beanspruchten, möglichst auszuschließen, wurde die Behandlungsdauer auf den Zeitraum von 3 bis 9 Wochen eingeschränkt. Durch die Beschränkung auf eine Mindestbehandlungsdauer von 3 Wochen sollte zudem verhindert werden, potentiell abgebrochene Behandlungen, die aus den Daten nicht direkt identifizierbar sind, auszuschließen. Die Definition des Behandlungsabbruchs wird in der Literatur uneinheitlich verwendet und umfasst zum Teil ein breites Spektrum an irregulären Entlassformen (vorzeitige Entlassungen auf ärztliche Veranlassung, mit und ohne ärztlichem Einverständnis, disziplinarische Entlassungen oder Entlassung aufgrund einer Verlegung) und eine entsprechend heterogene Patientengruppe (Lang, 2003). Einige Autoren konzeptualisieren Behandlungsabbrüche unabhängig von der Entlassform über die Behandlungsdauer. Dabei werden beispielsweise alle stationären Behandlungen, die innerhalb der ersten 14 Tage beendet werden, als Abbruch bewertet (Mussgay, Terporten & Rueddel, 1996). Andere Autoren definieren alle Entlassungen vor Ablauf von 20 Behandlungstagen als Behandlungsabbruch (Damke & Koechel, 2002). In der vorliegenden Arbeit wird mit dem Ausschluss aller Patienten, die kürzer als 3 Wochen behandelt werden, die Definition von Damke und Koechel aufgegriffen.

Ausschlusskriterien stellten lediglich das Alter und die Sprachkenntnisse dar. Patienten in einem Alter unter 18 Jahren wurden aufgrund der Beschränkung auf Aussagen über die Behandlung von Erwachsenen ausgeschlossen. Patienten deren sprachliche Kenntnisse nicht ausreichten, um die Fragebögen zu verstehen und zu beantworten, wurden ebenfalls nicht in die Stichprobe aufgenommen.

4.3 Operationalisierung der zu untersuchenden Merkmale

4.3.1 Operationalisierung des primären Outcomes und des Verlaufs

Das *primäre Outcome* dieser Untersuchung ist die Symptombelastung. Die Symptombelastung wurde zur Aufnahme, während des gesamten Verlaufs sowie zur Katamnese, also zu allen Messzeitpunkten, mit dem *Modul A (psychische und somatoforme Beschwerden)* der Kurzform der *Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH-49)*; Rabung, Harfst, Kawski, Koch, Wittchen & Schulz, 2009) erfasst. Der HEALTH Fragebogen wurde als Instrument zur generischen, multimodalen Erfassung zentraler Aspekte der psychosoziale Gesundheit entwickelt. Die Kurzform HEALTH-49 besteht aus sechs inhaltlich eigenständigen Modulen, die separat oder kombiniert eingesetzt werden können. Aus den sechs Modulen können neun Skalen (*Depressivität, Somatoforme Beschwerden, Phobische Angst, Psychische und Somatoforme Beschwerden, Psychisches Wohlbefinden, Interaktionelle Schwierigkeiten, Selbstwirksamkeit, Aktivität und Partizipation, Soziale Unterstützung und soziale Beeinträchtigung*) berechnet werden. In der vorliegenden Untersuchung wurde das Modul A *Psychische und somatoforme Beschwerden* mit 18 Items eingesetzt und anhand der gleichnamigen Skala ausgewertet. Die Skala *Psychische und somatoforme Beschwerden* (PSB) enthält 6 Items zur *Depressivität* (DEP), 7 Items zu *somatoformen Beschwerden* (SOM) und 5 Items zu *phobischen Ängsten* (PHO). Die jeweiligen Skalen werden durch Mittelwertsberechnung der Itemkennwerte gebildet. Für die Skala *Psychische und somatoforme Beschwerden* steht ein hoher Skalenwert für eine hohe Symptombelastung in den genannten Bereichen. Die Analysen der vorliegenden Arbeit wurden nur auf Gesamtskalenebene vorgenommen. Der HEALTH-49 bezieht sich auf das zurückliegende Zeitfenster von einer Woche und ist somit gut für das hier entwickelte Studiendesign, d.h. insbesondere für die wöchentlichen Verlaufsmessungen, geeignet.

Der HEALTH-49 wurde im Rahmen von zwei Untersuchungen (DETECT - Studie und QS-Reha® - Verfahren) psychometrisch überprüft (Rabung et al., 2009): Hinsichtlich der *Reliabilität* wurde die interne Konsistenz und die Trennschärfe der HEALTH-49-Skalen bzw. Items geprüft. Mit einem Cronbachs Alpha von 0.89 und einer mittleren Trennschärfe von 0.54 kann von einer sehr guten Messgenauigkeit der Skala *Psychische und somatoforme Beschwerden* ausgegangen werden. Die *konvergente Validität* des Moduls A wurde anhand der SCL-14 (Kurzform der Symptom Checklist-90, SCL-90-R; Franke, 2002; Harfst, Koch, Kurtz von Aschoff, Nutzinger, Rüdell & Schulz, 2002; Prinz, Nutzinger, Schulz, Petermann, Braukhaus & Andreas, 2008) überprüft: Es zeigen sich für alle drei Subskalen des Moduls A durchgängig hohe Korrelationen mit den äquivalenten Skalen der SCL-14 (SOM: $r = 0.85$; DEP: $r = 0.88$; PHO: $r = 0.84$). Die konvergente Validität der PSB Skala wurde anhand des

General Severity Index (GSI) der SCL-14 bzw. der SCL-90-R überprüft und zeigt mit einer Korrelation von 0.89 hohe Werte auf. Im Rahmen der DETECT-Studie konnte die *diskriminante Validität* anhand einer großen Hausarztstichprobe überprüft werden: Mit allen HEALTH-49-Skalen ist eine gute Differenzierung zwischen gesunden und psychisch erkrankten Menschen möglich. Eine besonders hohe Diskriminationsfähigkeit weist die Skala PSB auf (Area under the Curve (AUC) = 90.5). Von besonderer Relevanz für die vorliegenden Verlaufsmessungen ist die *Veränderungssensitivität* des Instruments. Zwar existieren bislang keine Daten zur Prüfung der Sensitivität hinsichtlich wöchentlicher Veränderungen, die *Veränderungssensitivität* anhand von standardisierten Prä-Post-Effektstärken der stationären Rehabilitationsbehandlung psychischer/psychosomatischer Erkrankungen stellt sich jedoch für alle Skalen als sehr zufriedenstellend dar (Prä-Post-Effektstärke der Skala PSB $d > 0.60$). Zudem ist durch die kurze Bearbeitungszeit und relativ geringen Fehlwertquoten eine gute Ökonomie und Praktikabilität gegeben (Rabung et al., 2009).

Für den HEALTH-49 existieren Referenzdaten sowohl für verschiedene Patientenpopulationen (stationäre Psychotherapiepatienten, Patienten mit psychischen und/oder körperlichen Erkrankungen aus der primärärztlichen Versorgung), als auch für eine gesunde Stichprobe (Rabung, Harfst, Koch & Schulz, 2008). Auf der Basis dieser Daten kann auf Parameter wie Mittelwert, Standardabweichung, Cut-Off-Wert oder Reliable Change Index der einzelnen Skalen zugegriffen werden.

4.3.2 Operationalisierung des sekundären Outcomes

Als *sekundäres Outcome* wurden *psychische Aspekte der gesundheitsbezogenen Lebensqualität/funktionalen Gesundheit* bei Aufnahme, bei Entlassung und zur 6-Monats-Katamnese mit der *Health Survey Short Form SF-8* (Ellert, Lampert & Ravens-Sieberer, 2005; Erhart, Wetzel, Krugel & Ravens-Sieberer, 2005; Ware, Kosinski, Dewey & Gandek, 2001) erhoben. Der SF-8 stellt eine Kurzform des Fragebogens zur gesundheitsbezogenen Lebensqualität SF-36 dar, eines der weltweit am häufigsten eingesetzten Instrumente zur Messung der gesundheitsbezogenen Lebensqualität (Bullinger, 2001; Bullinger & Kirchberger, 1998; Ware, 1993; Ware, 2000). Dabei handelt es sich um ein generisches, d.h. krankheitsübergreifendes Verfahren. Jedes der acht Items des SF-8 misst eine Dimension der subjektiv eingeschätzten Lebensqualität. Aus diesen acht Subskalen lassen sich zusätzlich die beiden Summenskalen der *Körperlichen* und *Psychischen Lebensqualität* bilden. Die Summenskala *Körperliche Lebensqualität* beinhaltet die folgenden vier Dimensionen: *Allgemeine Gesundheitswahrnehmung* (AG, General Health Perceptions), *Körperliche Funktionsfähigkeit* (KF, Physical Functioning), *Körperliche Rollenfunktionen* (KR, Role-Physical) sowie *Körperliche Schmerzen* (KS, Bodily Pain). Die Summenskala

Psychische Lebensqualität umfasst die vier Dimensionen *Vitalität* (VI, Vitality), *Soziale Funktionsfähigkeit* (SF, Social Functioning), *Psychisches Wohlbefinden* (PW, Mental Health) sowie *Emotionale Rollenfunktion* (ER, Role-Emotional). Eingeschätzt wird dabei der Zeitraum der zurückliegenden Woche. Die psychometrische Güte sowohl des SF-36 als auch des SF-8 konnte in zahlreichen Studien belegt werden (vgl. Bullinger, 2001; Bullinger & Kirchberger, 1998; Ware, 1993; Ware, 2000; Ware et al., 2001). Ein hoher Skalenwert weist auf eine hohe körperliche bzw. psychische gesundheitsbezogene Lebensqualität hin.

In der vorliegenden Studie soll die Summenskala der *Psychischen Lebensqualität* (Psychische Summenskala) als Maß für die psychischen Aspekte der gesundheitsbezogenen Lebensqualität verwendet werden. Entsprechend wird die Skala im Folgenden mit *SF-8 Psych.* abgekürzt. Für die deutsche Bevölkerung liegen repräsentative Normdaten für den SF-8 aus zwei Studien aus den Jahren 2003 und 2004 vor: Über eine telefonische Befragung durch das Robert-Koch-Institut aus dem Jahr 2003 konnten SF-8-Daten von insgesamt $N = 8\,318$ Personen erhoben werden (Ellert et al., 2005; Erhart et al., 2005). Im Rahmen einer schriftlichen Mehrthemenbefragung der Universität Leipzig aus dem Jahr 2004 wurden Daten von $N = 2\,552$ Personen erfasst (Beierlein, Bergelt, Bullinger, Brähler & Morfeld, 2010). Zum Vergleich der eigenen Ergebnisse mit den Normdaten des SF-8 werden aufgrund des Erhebungsverfahrens, d.h. der schriftlichen Befragung, und des Fragekontextes die Referenzwerte der Mehrthemenbefragung gewählt (Beierlein et al., 2010). Eine repräsentative Stichprobe psychisch erkrankter Patienten liegt für den SF-8 zurzeit noch nicht vor. Für Vergleiche der Ergebnisse der hier untersuchten Fragestellungen kann auf Daten einer großen Datenerhebung im Rahmen der Qualitätssicherung von 1 823 Patienten aus 11 Kliniken der Rehabilitation psychischer/psychosomatischer Erkrankungen zurückgegriffen werden (Schulz, Harfst, Andreas, Kawski, Koch & Rabung, 2009).

4.3.3 Operationalisierung der Prädiktormerkmale

4.3.3.1 *Soziodemographische und krankheitsbezogene Merkmale*

Bei der Aufnahme wurden folgende *soziodemographische Merkmale* durch Selbstangabe der Patienten erhoben: Alter, Geschlecht, Familienstand, Partnersituation, höchster Schulabschluss, höchster Berufsabschluss, Erwerbstätigkeit und monatliches Haushaltsnettoeinkommen. Zudem wurden klinische Patientenmerkmale über unterschiedliche Quellen erhoben: Die *Diagnosen* (alle vergebenen psychischen und somatischen Diagnosen pro Patient) wurden aus den Entlassberichten bzw. Arztbriefen der Kliniken erhoben. Über die Erfassung aller vergebenen Diagnosen eines Patienten können sowohl genauere Angaben zum Diagnosespektrum als auch zur Komorbidität gemacht werden. Als *krankheitsbezogene Variablen* wurden die Arbeitsunfähigkeitszeiten bzw. bei

nicht erwerbstätigen Patienten die Krankheitstage sowie die Inanspruchnahme stationärer und ambulanter psychotherapeutischer/psychiatrischer Vorbehandlungen in den letzten 6 Monaten vor Beginn der Behandlung anhand der Selbstangabe des Patienten erhoben.

4.3.3.2 *Personenbezogene Merkmale: Psychotherapiemotivation und Selbstwirksamkeitserwartung*

Psychotherapiemotivation

Als potentieller Prädiktor des Symptomverlaufs und des Behandlungsergebnisses wurde die *Psychotherapiemotivation* mit der Kurzform des *Fragebogens zur Psychotherapiemotivation (FPTM-23)*; Schulz et al., 2003; Schulz, Nübling & Rüdell, 1995) zu Behandlungsbeginn erhoben. Dieses Selbstbeurteilungsinstrument besteht aus 23 Items, die die sechs Skalen der Dimensionen *Psychischer Leidensdruck*, *Hoffnung*, *Verneinung psychischer Hilfsbedürftigkeit*, *Wissen*, *Initiative* und *Symptombezogene Zuwendung durch Andere* der Psychotherapiemotivation abdecken. Die Skalenwerte werden aus den Mittelwerten der jeweiligen Items gebildet. Ein hoher Skalenwert zeigt dabei jeweils eine starke Ausprägung der jeweiligen Dimension an. Die beiden Skalen *Verneinung psychischer Hilfsbedürftigkeit* und *Symptombezogene Zuwendung* sind negativ gepolt, sodass ein hoher Skalenwert hier eine geringe Psychotherapiemotivation anzeigt.

Der FPTM-23 weist gute psychometrische Kennwerte auf: Die Reliabilität (Cronbachs Alpha) liegt für die einzelnen Skalen zwischen .68 und .92. Die Konstruktvalidität wurde unter anderem mittels einer Fremdeinschätzung der Psychotherapiemotivation (5-stufiges Einzelitem) überprüft und erbracht für die vier Skalen *Leidensdruck*, *Hoffnung*, *Verneinung psychischer Hilfsbedürftigkeit* und *Initiative* signifikante Korrelationen; regressionsanalytisch konnten 49% der Varianz in der Fremdeinschätzung über die Skalen des FPTM-23 erklärt werden. Zusätzlich wurde die Konstruktvalidität an weiteren konstruktnahen Selbstbeurteilungsinstrumenten (gesundheitsbezogene Lebensqualität, Symptombelastung, Intention zur Mitarbeit und internale Kontrollüberzeugung) überprüft und kann als zufriedenstellend bewertet werden. Von besonderer Relevanz der hier untersuchten Prädiktionsfragestellung ist die prognostische Validität: Diese wurde anhand des Zusammenhangs mit dem Behandlungsergebnis bei stationärer psychotherapeutischer Behandlung eingeschätzt. Es zeigte sich, dass die Skalen *Leidensdruck* und *Hoffnung* die stärksten Zusammenhänge zum Ergebnis aufweisen, während die Skalen *Verneinung psychischer Hilfsbedürftigkeit*, *Initiative* und *Symptombezogene Zuwendung durch Andere* eine nur geringe prognostische Validität erreichten. Die Skala *Wissen* weist keine signifikanten Korrelationen mit dem Behandlungsergebnis auf (Nübling et al., 2006). Aufgrund der Kürze des Instrumentes ist die Praktikabilität gegeben (Nübling et al., 2006; Schulz et al., 2003).

Für die vorliegende Arbeit wurden aufgrund der dennoch geringen empirischen Basis zum Vorhersagewert hinsichtlich des Behandlungsergebnisses und aufgrund des explorativen Charakters der ersten Fragestellungen der Arbeit alle Skalen des FPTM-23 in der zweiten Fragestellung untersucht.

Selbstwirksamkeitserwartung

Als zweiter personenbezogener Prädiktor wurde die *Selbstwirksamkeitserwartung* untersucht. Dafür kam die deutsche Version der *Skala zur Allgemeinen Selbstwirksamkeitserwartung* (SWE, Schwarzer & Jerusalem, 1995) zur Anwendung. Die Skala umfasst 10 Items, die zu einem Summenscore addiert werden. Ein hoher Testwert steht dabei für eine hohe optimistische Kompetenzerwartung. Diese beschreibt das Vertrauen darauf, schwierige Situationen zu meistern, wobei der Erfolg auf die eigene Kompetenz zurück geführt wird. Die Skala ist ein weltweit verbreitetes Instrument und wurde in vielen Ländern hinsichtlich der Testgütekriterien untersucht. Die Reliabilität (Cronbachs Alpha) liegt in den deutschen Stichproben zwischen 0.80 und 0.90. Angaben zur diskriminanten und konvergenten Validität anhand der Korrelation zwischen SWE und anderen konstruktnahen sowie konstruktfernen Skalen finden sich z.B. bei Hinz, Schumacher, Albani, Schmid und Brähler (2006), die vor allem die Nähe zur Resilienz-Skala (Resilience Scale; Wagnild & Young, 1993), einem Instrument zur Erfassung der psychischen Widerstandsfähigkeit im Sinne einer Invulnerabilität, verdeutlichen ($r = 0.68$). Die Autoren stellen in ihrer Publikation bevölkerungsrepräsentative Normwerte einer deutschen Bevölkerungsstichprobe von $N = 2\,019$ Personen zur Verfügung.

4.3.3.3 Prozessmerkmale: Therapeutische Beziehung, Behandlungsart, -dauer und -dosis

Therapeutische Beziehung

Die *Güte der therapeutischen Beziehung*, die im Verlauf sowie zum Ende der Behandlung erhoben wurde, wurde über die deutsche Fassung (Bassler, Potratz & Krauthauser, 1995b) des *Helping Alliance Questionnaire* (HAQ; Luborsky, Barber, Siqueland, Johnson, Najavits, Frank et al., 1996) operationalisiert. Der Fragebogen besteht aus zwölf Items. Sechs der Items bilden die Subskala *Beziehungszufriedenheit*, fünf Items erfassen die *Erfolgswilligkeit* und ein Item bildet die globale Erfolgseinschätzung im Vergleich zum Behandlungsbeginn ab. Zusätzlich zur patientenseitig auszufüllenden Version existiert auch eine in der dritten Person formulierte Version, die von dem Therapeuten ausgefüllt wird. In der vorliegenden Untersuchung wird die therapeutische Beziehung ausschließlich über die Skala *Beziehungszufriedenheit* aus Patientenperspektive erhoben.

Untersuchungen der Testgütekriterien der Originalfassung zeigen für beide Skalen eine hohe Reliabilität (Cronbachs Alpha sowie Retestreliabilität) sowie eine gute Konstruktvalidität

(Luborsky et al., 1996). Die deutsche Version des HAQ wurde in einer Studie mit 239 stationär behandelten Patienten mit psychischen Erkrankungen (hauptsächlich ICD-10-Diagnosen F3 - F6) hinsichtlich der Testgütekriterien überprüft (Bassler et al., 1995b). Dabei wurde die Güte der therapeutischen Beziehung erstmals 2 Wochen nach Behandlungsbeginn erfasst, um sicherzustellen, dass eine Einschätzung der Beziehung bereits möglich ist. In den Studien, die die Arbeitsgruppe um Luborsky im ambulanten Setting durchführte, wurde die Güte der therapeutischen Beziehung in der Regel nach der dritten Behandlungswoche eingeschätzt (Luborsky et al., 1996). Durch die im stationären Setting vorzufindende erhöhte Dichte psychotherapeutischer Behandlungsmaßnahmen (im Sinne einer höheren Dosis in einem bestimmten Zeitraum) ist eine frühere Erhebung der therapeutischen Beziehung zu rechtfertigen. Entsprechend wird für die vorliegende Untersuchung im stationären Setting als Erhebungszeitpunkt die zweite Behandlungswoche gewählt. Bassler et al. (1995b) fanden eine hohe *Reliabilität* (Cronbachs Alpha: 0.89). Ebenfalls konnten die Autoren zeigen, dass die Güte der eingeschätzten Beziehung positiv mit dem Therapieerfolg zusammenhängt. Auch hinsichtlich der *konvergenten Validität* ergeben sich zufriedenstellende Ergebnisse. Aufgrund der Kürze des Instruments und der hohen *Akzeptanz* bei Patienten und Therapeuten ist der HAQ gut in der klinischen Routine einsetzbar (Bassler et al., 1995b).

Behandlungsdauer, -dosis und -art

Die *Dauer* der stationären Behandlung wurde über die Differenz zwischen Aufnahme- und Entlassdatum gebildet, die aus der Klinikdokumentation bzw. den Entlassberichten und Arztbriefen entnommen werden konnten.

Die *Behandlungsdosis* wurde über die Anzahl und Dauer der erhaltenen Behandlungsmaßnahmen durch die *Klassifikation Therapeutischer Leistungen (KTL)*; Deutsche Rentenversicherung Bund, 2007) erhoben. Die KTL ist ein Leistungskodiersystem der Deutschen Rentenversicherung, mit dem alle abrechnungsrelevanten Behandlungen anhand der KTL-Kodierung im Entlassbericht erfasst und in Listenform mit der Angabe der Anzahl und der Dauer jeder vorgehaltenen Maßnahme für jeden Patienten dokumentiert werden. Dieses Instrument wird in der Routinedokumentation in allen Einrichtungen, deren Träger die Deutsche Rentenversicherung ist, eingesetzt. Die Behandlungsmaßnahmen von Patienten, deren Behandlung nicht über die Deutsche Rentenversicherung abgerechnet wird, werden im Regelfall nicht mittels KTL-Kodierungen dokumentiert. In einer der drei untersuchten Kliniken wurden hauptsächlich Akutpatienten behandelt, deren Behandlung nicht über die Deutsche Rentenversicherung finanziert wird. In dieser Klinik wurden die Behandlungen jedoch routinemäßig für alle Patienten über die KTL dokumentiert. Somit

konnte die Erfassung von Dosisvariablen für einen Großteil der Stichprobe gewährleistet werden.

Von besonderem Interesse war in der vorliegenden Untersuchung die Frage nach der Dosis psychotherapeutischer Verfahren. In dem multimodalen Behandlungskonzept des untersuchten Settings sollte daher eine Abgrenzung der Dosis von der Vielzahl eher psychotherapieferner Maßnahmen vorgenommen werden. Der psychotherapieferne Bereich umfasst dabei beispielsweise Sport- und Bewegungstherapien, Physiotherapie oder Psychoedukation. Die Einordnung der Kategorien ist nicht unumstritten und kann kritisch diskutiert werden. Ein Merkmal zur Einordnung in psychotherapienahe Maßnahmen bildete die Auswahl hinsichtlich der Berufsgruppen, die die Maßnahmen durchführen. Als psychotherapienahe Maßnahmen werden demnach die Maßnahmen bezeichnet, die von Therapeuten mit einer laufenden oder abgeschlossenen Psychotherapieausbildung (nach dem Psychotherapeutengesetz, PsychThG; FNA: 2122-5) durchgeführt werden. Dieses Kriterium diene der Genauigkeit der resultierenden Aussagen, die über die Dosis psychotherapienaher Maßnahmen gemacht werden. Die Kategorie *Psychotherapienahe Maßnahmen* wurde in Anlehnung an die von Barghaan et al. durchgeführte Einteilung (Barghaan et al., 2007; Barghaan et al., 2005) vorgenommen. Diese Einteilung stellte auch die Grundlage für die Entwicklung von Leitlinien der rehabilitativen Behandlung depressiver Erkrankungen dar (Barghaan, Schulz, Steinmann, Krattenbacher, Koch & Dirmaier, 2010; Dirmaier, Krattenmacher, Watzke, Koch, Schulz & Barghaan, 2010). In der Kategorie *psychotherapienah* werden die folgenden Maßnahmen aus der KTL 2007 (Deutsche Rentenversicherung Bund, 2007) subsumiert: F01-F069, F08-F099, G02-G099, G11-G203. Zur Untersuchung der *Art der erhaltenen Psychotherapiemaßnahmen* wurde ebenfalls in Anlehnung an Barghaan et al. eine Einteilung in *psychodynamische psychotherapeutische Behandlungsmaßnahmen* (PD-T; KTL-Kodierungen: G03-G049, G08-G089, F021) und *verhaltenstherapeutische Behandlungsmaßnahmen* (VT; KTL-Kodierungen: G05-G079, F022) vorgenommen. Dabei wurde entsprechend des Vorgehens bei Barghaan et al. die Minutenzahl der einzelnen Maßnahmen aufsummiert und anschließend das prozentuale Verhältnis verhaltenstherapeutischer Maßnahmen zur Gesamtsumme aller verhaltenstherapeutischen und psychodynamischen Maßnahmen berechnet. Äquivalent dazu wird der Anteil an psychodynamischen Behandlungsmaßnahmen, gemessen an der Gesamtpsychotherapie, berechnet (Barghaan et al., 2007; Barghaan et al., 2005).

4.3.4 Operationalisierung der theoriegeleiteten Verlaufsgruppen

Für die Bildung der *theoriegeleiteten Verlaufsgruppen* muss ein zeitliches Kriterium (zu welchem Zeitpunkt während der Behandlung wird eine relevante Verbesserung erwartet?) und ein qualitatives Kriterium (was ist eine relevante Verbesserung?) definiert werden. In der

vorliegenden Arbeit wurden als qualitative Kriterien Differenzwerte und Effektstärken der Veränderungen herangezogen. Es wurden drei unterschiedliche Verlaufgruppen bestimmt: 1. Frühe Verbesserungen (*Early Response*) wurden über die Differenz zwischen der Symptombelastung bei Behandlungsbeginn (Prä-Messung) und der Symptombelastung innerhalb der ersten 3 Wochen des stationären Aufenthalts definiert. Aufgrund der Verlaufsmessungen besteht hier die Möglichkeit, das *zeitliche Kriterium* einer Early Response nicht alleine auf einen Messzeitpunkt festzulegen, sondern über einen Zeitraum mehrerer Messungen hinweg zu bilden. Hier wurden die Messzeitpunkte nach der 2. und nach der 3. Woche gewählt. Da die bisherige empirische und theoretische Festlegung des kritischen Early Response-Zeitpunkts insbesondere für das stationäre Setting unklar ist, stellt diese Vorgehensweise eine erweiterte Möglichkeit dar, Patienten mit frühem Ansprechen auf die Behandlung zu identifizieren. Die Einschränkung auf die ersten 3 Wochen wurde gewählt, da bis zu diesem Zeitpunkt klinisch relevante Veränderungen erwartet werden können (vgl. Abschnitt 2.5). Die erste Woche wurde nicht eingeschlossen, da zu diesem Zeitpunkt unspezifische („Erholungs“-) Effekte auftreten können, die von Effekten der psychotherapeutischen Behandlung nicht klar abgegrenzt werden können. Patienten, die in die Gruppe Early Response fallen, müssen demnach nach der zweiten *oder* dritten Woche eine relevante Verbesserung aufweisen. Für das *qualitative Kriterium* einer relevanten Verbesserung wurde für jeden Patienten nach dem Konzept der statistischen Signifikanz überprüft, ob die erreichte Verbesserung statistisch reliabel ist, berechnet über den Reliable Change Index (RCI; Jacobson & Truax, 1991; für eine differenzierte Beschreibung s. Abschnitt 4.5.4). Das Kriterium einer Response wurde darüber hinaus über das Ausmaß der Effektstärke der Verbesserung definiert. Dabei wurde die jeweilige Differenz zum Ausgangswert der Belastung durch die Standardabweichung der Ausgangsbelastung geteilt. Die notwendige Höhe der Effektstärke wurde bei mindestens mittlerer Effektstärke festgelegt, entsprechend der Konvention von Cohen für die Interpretation der Effektstärken (Cohens d mit den Grenzen $d > 0.20$ für kleine, $d > 0.50$ für mittlere und $d > 0.80$ für große Effekte; Cohen, 1988; vgl. Abschnitt 4.5.4). Eine Response ist demnach dann gegeben, wenn die Verbesserung von Aufnahme bis zu einem kritischen Zeitpunkt mindestens die kritische Differenz des RCI *sowie* mindestens die Untergrenze einer mittleren Effektstärke d überschreitet.

Die zweite Gruppe wurde als *Late Response* bezeichnet. Diese Gruppe der verzögerten Verbesserungen wurde über das Erreichen des qualitativen Response-Kriteriums zum Entlasszeitpunkt definiert und über die Differenz von Prä-Post-Werten operationalisiert. Eine Late Response ist demnach gegeben, wenn ein Patient das Kriterium der Response *nicht* innerhalb der ersten 3 Wochen, aber zum Entlasszeitpunkt erreicht. Wurde das Response-Kriterium weder während der ersten 3 Wochen, noch zum Entlasszeitpunkt erreicht, so

wurde dieser Patient der Gruppe der *Non-Response* zugeordnet. Patienten, die sich im Behandlungsverlauf verschlechterten, wurden in der Gruppe Non-Response subsumiert. Die Gruppe der verschlechterten Patienten ist eine besondere Gruppe, für die, wegen der zu erwartenden geringen Fallzahlen (ca. 5%), gesonderte Analysen nur dann vorgenommen werden, wenn in diese Gruppe mindestens 50 Patienten fallen.

Die Zuordnung zu der theoriegeleiteten Gruppe Early Response ist unabhängig von dem weiteren Verlauf. Das bedeutet, dass Patienten, die sich sowohl nach der 2. oder 3. Woche, als auch zur Entlassung nach dem festgelegten Kriterium verbessert haben der Gruppe der Early Response zugeordnet bleiben. Für den Fall, dass ein Patient keinen der beiden Fragebögen der 2. und 3. Woche ausgefüllt hat (Missing), kann dieser Patient nicht der Gruppe der Early Response zugeordnet werden. Da die Late Response in dieser Definition von dem Nichtvorhandensein einer Early Response abhängt, wurde bei einer nicht bestimmbarer Zuordnung nach der 2. oder 3. Woche keine weitere Zuordnung zu den anderen inhaltlichen Gruppen vorgenommen. Alle Patienten, die aufgrund entsprechender fehlender Werte nicht zugeordnet werden konnten, wurden in der Gruppe *n/a* (Abkürzung für *nicht anwendbar*) zusammengefasst. Tabelle 1 zeigt eine schematische Darstellung der Bildung der theoriegeleiteten Verlaufsgruppen.

Tabelle 1: Schematische Darstellung der theoriegeleiteten Gruppenbildung

		Effektstärke bei Entlassung		
		> 0.50	< 0.50	Missing
Effektstärke nach Woche 2 oder 3	> 0.50	Early Response	Early Response	Early Response
	< 0.50	Late Response	Non-Response	n/a
	Missing ¹	n/a ²	n/a	n/a

¹ fehlender Wert

² nicht anwendbar: Restkategorie, für die eine Bestimmung der Responseform aufgrund fehlender Werte nicht möglich ist.

4.4 Datenmanagement

4.4.1 Aufbereitung der Daten

Für die vorliegende Untersuchung wurden Daten aus verschiedenen Datenquellen ausgewertet. Zum einen wurden die Patientenangaben per Hand in eine Datenmaske

eingetragen. Daten aus den Entlassberichten sowie Arztbriefen konnten anhand der pseudonymisierten Kennungen der Patienten über fortlaufende Ziffern den Patientendaten zugeordnet werden. Für die Repräsentativitätsanalysen lagen zum anderen zusätzlich Daten aus den Klinikdokumentationssystemen vor. Diese wurden in elektronischer Form von den Kliniken übermittelt. Über die durchgängige Kodierung konnten die Patientenangaben mit den Daten aus der Klinikdokumentation verknüpft und zusammengefügt werden.

4.4.2 Umgang mit fehlenden Werten

Mithilfe von Fehlwertanalysen kann die Struktur des Datenausfalls aufgedeckt und hinsichtlich systematischer Effekte untersucht werden. Nach der Analyse der fehlenden Werte wurden Entscheidungen über die Anwendung von Ersetzungsverfahren getroffen.

4.4.2.1 Analyse fehlender Werte

Zur Analyse der Struktur der fehlenden Werte in dem HEALTH-49 Modul A *Psychische und somatoforme Beschwerden* (primäres Outcome) wurde zunächst die Systematik und das Ausmaß des Datenausfalls analysiert. Generell sind die fehlenden Werte pro Variable über alle Fälle hinweg und die fehlenden Werte pro Fall über (mehrere oder alle) Variablen hinweg zu unterscheiden. Da hier gezielt die Struktur und Ersetzung der fehlenden Werte der Outcome-Maße von Interesse sind, bezieht sich die Angabe „pro Fall“ hier genau genommen nur auf die Items des jeweiligen Outcome-Instruments und nicht auf den gesamten Datensatz eines Falles bzw. einer Person. Als Richtlinie gilt, dass mit dem Einschluss von Fällen oder Variablen, die mehr als 30% fehlende Werte aufweisen, mehr Unsicherheiten und Fehler in Kauf genommen werden, als Informationen gewonnen werden können (Wirtz, 2004). Diese Fälle bzw. Variablen liefern demnach ungenaue Schätzungen und sollten von den weiteren Analysen ausgeschlossen bleiben. Bei jedem Messzeitpunkt tauchen fehlende Werte auf, deren Anzahl pro Person variiert. Hinsichtlich des Richtwerts wurden die Daten daraufhin geprüft, wie viele Personen (pro Messzeitpunkt) mehr als 30% fehlende Werte in den 18 Items des HEALTH-49 A Moduls aufweisen. Die Anzahl der Personen, die mehr als 30% fehlende Werte aufweisen, d.h. mehr als 5 Items nicht ausgefüllt haben, variiert zwischen den Messzeitpunkten von 0 bis 6 und beträgt maximal 1% der Stichprobe zu dem jeweiligen Messzeitpunkt. Diese Fälle werden von den weiteren Analysen ausgeschlossen. Hinsichtlich der Variablen wurden alle 18 Items des HEALTH-49 A von mindestens 70% der Personen zu dem jeweiligen Messzeitpunkt ausgefüllt.

Zur Bestimmung der Anzahl der fehlenden Werte, für die eine Ersetzungsmethode zur Anwendung kommen soll (< 30% fehlender Werte pro Fall), wurde für jeden der 12

Messzeitpunkte (t0, w1 - w9, t1, t2) die maximale Anzahl möglicher vorhandener Werte wie folgt berechnet,

$$N_{\max \text{ zu Tx}} = N_{\text{Person zu Tx}} * N_{\text{Variablen zu Tx}}$$

wobei $N_{\text{Person zu Tx}}$ über die Messzeitpunkte variiert ($N = 21$ bis $N = 790$) und $N_{\text{Variablen zu Tx}}$ (Anzahl der Variablen zu Zeitpunkt x) immer 18 beträgt (18 Items des HEALTH-49 A Moduls). $N_{\max \text{ zu Tx}}$ variiert von 378 (Woche 9) bis 14 220 (t0). Die Summe aller möglichen vorhandenen Werte über alle Messzeitpunkte hinweg beträgt $N_{\max \text{ gesamt}} = 83\,034$ Datenpunkte. Die Gesamtzahl fehlender Datenpunkte über alle (eingeschlossenen) Personen und Messzeitpunkte hinweg beträgt $n = 371$ und macht somit lediglich 0.5% aller möglichen Werte aus. Die prozentuale Anzahl der zu ersetzenden Fehlwerte variiert zwischen den Messzeitpunkten von 0 – 0.6% und liegt damit durchgängig in einem sehr niedrigen Bereich. Zusammenfassend zeigt sich in den vorliegenden Daten eine geringe Anzahl an fehlenden Werten im HEALTH-49 A, der Anteil der Daten, die für dieses Instrument ersetzt werden, beträgt 0.5%.

Für alle anderen psychometrischen Instrumente (SF-8, FPTM-23, SWE, HAQ) zeigt die Fehlwertanalyse auf Variablenebene, dass in keinem der Items aller Instrumente jeweils mehr als 2% fehlende Werte auftreten. Fallbezogen wurde ebenfalls für jedes Instrument bzw. die zu berechnende Skala Personen aus den weiteren Analysen ausgeschlossen, die mehr als 30% fehlende Werte aufwiesen.

Hinsichtlich der Prozessmerkmale, die auf den über die KTL Daten ermittelten Dosisangaben gebildet wurden, treten keine variablenbezogenen Fehlwerte auf. Da nicht für alle Patienten ein Entlassbericht vorlag, sind für einen bestimmten Anteil der Patienten keine Dosisangaben vorhanden (vgl. Abschnitt 4.3.3.3).

4.4.2.2 Ersetzung fehlender Werte

Zum Umgang mit fehlenden Werten steht eine Vielzahl methodischer Vorgehensweise zur Verfügung. Neben dem Ausschluss von fehlenden Werten oder ganzen Fällen mit fehlenden Werten (fallweiser oder paarweiser Ausschluss), die als Eliminierungsverfahren bezeichnet werden, stehen die sogenannten Imputationsverfahren, bei denen die fehlenden Werte durch geschätzte Werte auf der Basis der vorhandenen Werte der Stichprobe ersetzt werden, zur Verfügung (für einen Überblick zu Ersetzungsverfahren vgl. Lüdtkke, Robitzsch, Trautwein & Köller, 2007; Rabung, 2007). Eine Ersetzungsmethode, bei der alle verfügbaren Informationen des Datensatzes zur Ersetzung einbezogen werden können, und die als weitgehend stabil, d.h. relativ unabhängig von dem Ausfallmuster der fehlenden Werte in der Datenmatrix gilt, stellt der Expectation Maximization Algorithmus (EM Methode) dar (Rabung,

2007; Rabung, 2010). Das Verfahren basiert auf Maximum-Likelihood-Schätzungen und folgt einem iterativen Prozess. Bei jeder Iteration werden in einem ersten Schritt (Expectation-Step) die fehlenden Werte auf der Basis einer Log Likelihood Funktion geschätzt, in einem zweiten Schritt (Maximization-Step) werden aus den vollständigen Daten Parameterschätzer berechnet, die in der folgenden Iteration wiederum zur erneuten Schätzung der fehlenden Werte herangezogen werden. Der iterative Prozess wird solange durchgeführt, bis die Parameter der Iterationen sich nur noch minimal unterscheiden. Der Vorteil dieser Methode besteht darin, dass fehlende Werte prinzipiell aus allen anderen vorhandene Daten des Falles geschätzt werden können. Außerdem ist dieses Verfahren weniger an restriktive Annahmen hinsichtlich des Ausfallmusters der Fehlwerte gebunden (für eine detaillierte Beschreibung des Verfahrens vgl. Little & Rubin, 2002; Rabung, 2007).

Zur Ersetzung der fehlenden Werte der HEALTH-49 A Skala wurde der EM-Algorithmus angewendet. Dabei wurden alle anderen psychometrischen Skalen in die Schätzfunktion einbezogen. Keinen Eingang in die Ersetzung fanden dagegen soziodemographische Angaben und Angaben, die nicht von den Patienten selbst gemacht wurden (beispielsweise die Diagnosen).

Missings in den anderen psychometrischen Skalen (SF-8, FPTM-23, SWE) wurden aufgrund der geringen Gesamtanzahl nicht ersetzt. Die fehlenden Werte aus der zweiten Behandlungswoche im HAQ (Güte der therapeutischen Beziehung) wurden durch die Daten aus der dritten Woche ersetzt. Für alle anderen Daten wie soziodemographische und klinische Patientenmerkmale oder Angaben zum Behandlungsprozess wurden aus inhaltlichen Gründen keine Ersetzungsverfahren angewendet. Dadurch ergaben sich in den Unterschiedstestungen hinsichtlich der Patienten- und Prozessmerkmale jeweils geringfügig unterschiedliche Stichprobengrößen. Die auf den KTL-Daten basierenden Analysen zur Behandlungsdosis und -art wurden ebenfalls entweder mit einem reduzierten Datensatz durchgeführt oder, bei Verursachung eines zu hohen Datenverlustes für multivariate Analysen, ggf. aus den Analysen ausgeschlossen. Die für verschiedene Analysen geringfügig variierenden Stichprobengrößen sind in jeder Ergebnistabelle angegeben.

4.5 Statistische Auswertungsmodelle

4.5.1 Stichprobenbeschreibung und Drop-Out-Analysen

Zur Beschreibung der Verlaufsstichprobe und der Verlaufs-Outcome-Stichprobe werden Mittelwerte und Standardabweichungen der kontinuierlichen Merkmale angegeben, für kategoriale Merkmale erfolgt die Darstellung der Häufigkeitsverteilungen. Zur Untersuchung von systematischen Selektionseffekten hinsichtlich der Patienten, die die Studienteilnahme während der Behandlung abgebrochen haben (Drop-Out-Stichproben) werden diese in

mehreren relevanten Merkmalen auf signifikante Unterschiede zur Analysestichprobe getestet. In Abhängigkeit von dem Skalenniveau kommen dazu t -Tests und χ^2 -Tests, bei geringen Zellbesetzungen (wenn der Anteil der erwarteten Zelhäufigkeiten, die kleiner als 5 sind, 20% überschreiten) Fishers exakte Tests (*FE*), zur Anwendung (Bortz, 2005). Zur Einordnung der von der Untersuchungsstichprobe erreichten Werte in den psychometrischen Skalen HEALTH-49 A, SF-8 Psych. und SWE werden Vergleiche mit Referenzgruppen oder bevölkerungsrepräsentativen Normstichproben mittels Ein-Stichproben- t -Test durchgeführt. Dabei wird für die jeweilige Skala geprüft, ob sich die Mittelwerte der Stichprobe signifikant von denen der Referenzgruppen unterscheiden.

4.5.2 Verlaufsanalyse

4.5.2.1 Explorative Verlaufsanalyse: Growth Mixture Modeling

Zur Identifikation von latenten Symptomverlaufsklassen wurde die Methode der Wachstumskurvenmodellierung angewendet. Dabei werden längsschnittliche Verlaufspfade von Individuen modelliert, um Subgruppen mit möglichst homogenen Verläufen zu identifizieren.

Konventionelle Wachstumsmodellierungen, d.h. die statistische Beschreibung von längsschnittlichen Verlaufsdaten, setzen voraus, dass alle untersuchten Individuen einer in Bezug auf das Veränderungsmerkmal homogenen Population entstammen und somit eine angemessene Schätzung der Gesamtpopulation mit einer einzigen Wachstumskurve (Trajektorie) gegeben ist (Eid et al., 2008; Jung & Wickrama, 2008). Ebenso wird hier der Einfluss von Kovariaten auf die Wachstumsparameter für alle Individuen in gleicher Weise vorausgesetzt. Theoretisch-konzeptuelle Erwägungen und empirische Befunde deuten jedoch darauf hin, dass im Verlauf von Psychotherapie von unterschiedlichen Patientensubpopulationen mit unterschiedlichen prototypischen Verläufen ausgegangen werden kann (s. Abschnitt 2.5.7). Bei der Schätzung einer einzigen Wachstumskurve könnten Unterschiede in sich bedeutsam unterscheidenden Subgruppen über die „Mittelung“ der Daten verloren gehen, bzw. wichtige interindividuelle Unterschiede nicht entdeckt werden. Um diese latente Heterogenität innerhalb der Population berücksichtigen zu können, wurden Wachstumskurvenmodelle zu Mischverteilungsmodellen weiterentwickelt (Muthén, 2004).

In der vorliegenden Arbeit wird Heterogenität innerhalb der Gesamtpopulation hinsichtlich der Symptomverläufe angenommen, die auf nicht beobachtete Subpopulationen zurückzuführen ist. Um diese latenten Symptomverlaufsklassen zu identifizieren, soll ein neuer Ansatz der Wachstumsmodellierung, das sogenannte *Growth Mixture Modeling* (*GMM*), auch als Mischverteilungs-Wachstumskurven-Modell benannt, angewendet werden.

Die Methode des GMM basiert darauf, dass latente Klassen innerhalb der Population angenommen und identifiziert werden können, d.h. es wird getestet, ob verschiedene Klassen mit prototypischen Verlaufsmustern existieren. Über mehrere Messzeitpunkte werden dabei aus den Kurvenparametern jedes einzelnen Patienten Klassen von Verlaufskurven geschätzt, innerhalb derer die Verlaufparameter der zu einer Klasse zugehörigen Patienten möglichst homogen und die Parameter der Patienten aus unterschiedliche Klassen möglichst heterogen sein sollen. Falls solche Klassen identifizierbar sind, kann mit Hilfe des GMM für jedes Individuum die Wahrscheinlichkeit für die Zugehörigkeit zu jeder der latenten Klassen geschätzt werden.

Bei der GMM-Methode handelt es sich um ein exploratives Vorgehen, bei dem weder die Anzahl der Klassen, noch die Anzahl der zu den Klassen zugeordneten Individuen vorab festgelegt wird. In einem schrittweisen Verfahren werden sukzessiv aufsteigende Klassenlösungen modelliert, die anhand von Informationskriterien und Modellgütekriterien hinsichtlich des Ausmaßes der Modellpassung miteinander verglichen werden können. Für die Entscheidung über die beste (d.h. die am besten die Daten beschreibende) Klassenlösung wird empfohlen, zu den statistischen Kennwerten auch inhaltliche Gesichtspunkte der spezifischen Untersuchung (Fragestellung, Design und Setting etc.) einzubeziehen. Für die gefundenen k Klassen resultiert jeweils eine gemittelte Wachstumstrajektorie.

Die Kurvenparameter, die in die Analysen der vorliegenden Arbeit einbezogen wurden, sind der *Intercept* (Achsenabstand), der *lineare* und der *quadratische Slope* (Steigungskoeffizient). Der Intercept bildet die Merkmalsausprägung, hier die Symptombelastung, zum ersten Messzeitpunkt ab. Der *lineare Slope* beschreibt die Steigung der Kurve über den Gesamtverlauf aller vorhandenen Messzeitpunkte und somit die lineare Veränderung des Merkmals über die Zeit. Um zu den linearen Verläufen auch log-lineare Verläufe abbilden zu können, wird der *quadratische Slope*, die Beschleunigung bzw. Verlangsamung der Veränderung, einbezogen. Für die Berechnung des linearen und des quadratischen Slopes einer Verlaufskurve sind Daten von mindestens drei Messzeitpunkten erforderlich. Abbildung 2 veranschaulicht exemplarisch die drei Parameter der Verlaufskurven.

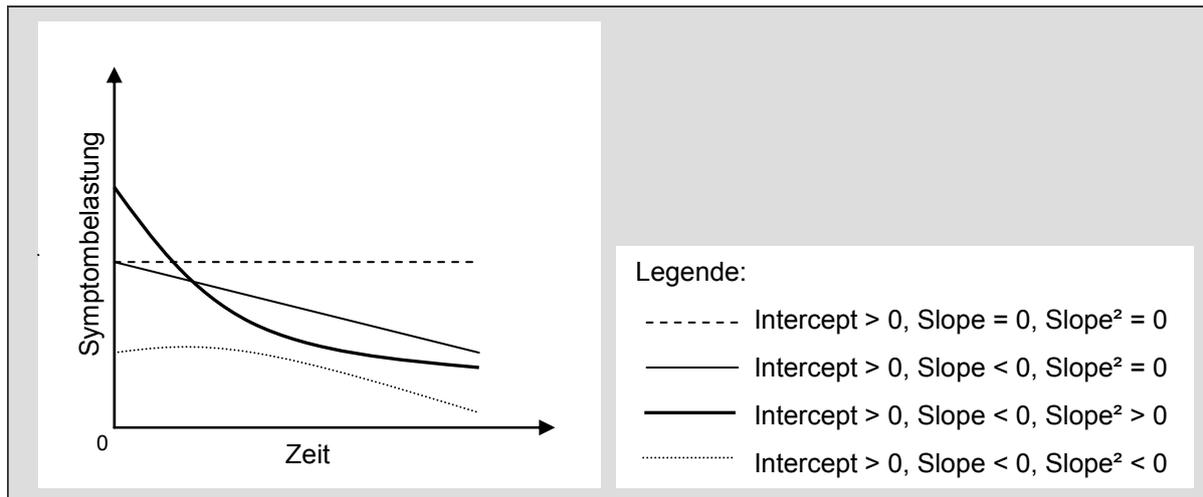


Abbildung 2: Exemplarische Verlaufskurven für unterschiedliche Ausprägungen der Kurvenparameter Intercept, linearer und quadratischer Slope

4.5.2.2 Statistische Kriterien des Growth Mixture Modeling

Für jede identifizierte Klassenlösung werden mehrere Kennwerte der Modellpassung berechnet. Zur Entscheidung über die Klassenlösung mit der besten Modellpassung werden mehrere statistische Kennwerte, Informationskriterien und Ergebnisse inferenzstatistischer Testungen herangezogen. Die Modellparameter werden über eine Maximum-Likelihood-Schätzfunktion bestimmt. Häufig werden das *Bayesian Information Criterion (BIC)* und das *Akaike Information Criterion (AIC)* angegeben. Diese Kriterien beschreiben die Güte der Modellpassung, wobei der jeweils niedrigste Wert mit der besten Modellpassung der Daten einhergeht. Bei der Berechnung des BIC wird im Gegensatz zum AIC die Stichprobengröße berücksichtigt. Zusätzlich kann inferenzstatistisch mithilfe des *Lo-Mendell-Rubin Likelihood Ratio Tests (LMR-LRT)* (Lo, Mendell & Rubin, 2001) überprüft werden, ob die Klassenlösung mit k Klassen die Daten signifikant besser schätzt als die Klassenlösung mit $k - 1$ Klassen. Bei der schrittweise vorwärts gerichteten Klassentestung würde eine Klassenlösung mit k Klassen, die sich nicht mehr signifikant von der Klassenlösung mit $k - 1$ Klassen unterscheidet, als Gütekriterium für die Lösung mit $k - 1$ Klassen gelten. Zur Signifikanzprüfung bei kleinen Stichproben (n pro k Klassen) stellt der *Bootstrap Likelihood Ratio Test (BLRT)* das robustere Verfahren dar. Die Bootstrap-Methode (Efron, 1979) basiert auf der Iteration von Zufallsstichprobenziehungen aus den vorhandenen Daten, um die unbekannte theoretische Verteilung der Daten empirisch zu überprüfen. Durch die Potenzierung der empirischen Kennwerte, die durch das n -fach wiederholte Ziehen mit Zurücklegen ermöglicht wird (maximale Anzahl verschiedener Stichproben: n^n), können mittels Bootstrapping robustere Schätzungen für kleine Stichproben erreicht werden.

Ein weiteres Kriterium zur Einschätzung der Modellgüte zeigt die *Log Likelihood (LL)* an. Die Parameter werden so lange adjustiert, bis sich die Log Likelihood nicht mehr weiter erhöht

bzw. bis die Anpassung nur noch zu unbedeutenden Veränderungen in den Parametern führt. Kann der Log Likelihood Wert repliziert werden, deutet dies auf die statistische Sicherheit des Modells hin.

In Anbetracht der aktuell geführten Diskussion zu der Güte der einzelnen Modellfit-Kriterien wird hinsichtlich der Informationskriterien das BIC als zuverlässiger Kennwert und für die Signifikanztestung der unterschiedlichen Klassenlösungen der BLRT als robust eingeschätzt (Jung & Wickrama, 2008). Einigkeit herrscht darüber, zur Entscheidung über die Klassenlösung nicht nur die statistischen Kennwerte, sondern immer auch inhaltliche Erwägungen einzubeziehen.

4.5.2.3 *Inhaltliche Kriterien des Growth Mixture Modeling*

Neben den Statistischen Parametern werden auch inhaltliche Kriterien bei der Entscheidung über die Lösung mit dem besten Modellfit herangezogen. Zum einen wird die minimale Klassengröße inhaltlich festgelegt. Da es im Rahmen der Analysen um die Identifikation prototypischer Verläufe und nicht um die Beschreibung von Einzelfällen bzw. Extremverläufen, die nur selten vorkommen, geht, wird die Mindestgröße einer Klasse auf 5% der Gesamtstichprobe festgelegt. Ein zweites Kriterium berücksichtigt die Genauigkeit der Klassenzuordnung. Neben der Entropie, die die mittlere höchste Klassenwahrscheinlichkeit aller Patienten einer Klasse anzeigt, kann dieser Kennwert der Klassenzuordnungswahrscheinlichkeit für jeden Patienten einzeln betrachtet werden. Sollte die maximal erreichte Klassenwahrscheinlichkeit bei einem Patienten die 50% nicht erreichen, so ist die Zuordnung mit einer zu geringen Genauigkeit geschätzt. Dieser Patient wird nachträglich aus der Klasse ausgeschlossen, um in jeder Klasse ausschließlich Patienten mit einer Zuordnungsgenauigkeit von mindestens 50% zu haben. Als drittes inhaltliches Kriterium sollen gegebenenfalls Überlegungen zur Differenziertheit, zum Informationsgewinn sowie zur Interpretierbarkeit der verschiedenen Klassenlösungen gegeneinander abgewogen werden. Für die statistisch identifizierten und inhaltlich modifizierten Verlaufgruppen werden zusätzlich die Konfidenzintervalle der relativen Häufigkeiten der einzelnen Gruppen berechnet.

Insgesamt lässt sich festhalten, dass mit dem GMM-Ansatz eine genauere Beschreibung des therapeutischen Verlaufs ermöglicht wird, die über die Annahme einer einzigen Verlaufskurve für alle Patienten hinausgeht (Lutz et al., 2007b). Ein weiterer Vorteil gegenüber clusteranalytischen Verfahren ist die Robustheit des Verfahrens im Umgang mit fehlenden Werten und unterschiedlich stark besetzten Messzeitpunkten. Die Vorteile dieser Modelle für die Untersuchung von Therapieverläufen konnten in einigen Studien aufgezeigt werden (Keller & Hautzinger, 2007; Lutz et al., 2007a).

Die GMM-Analysen zur Identifikation von Symptomverlaufsklassen wurde mit der Statistiksoftware Mplus Version 5.1 (Muthén & Muthén, 2008) durchgeführt.

4.5.2.4 Theoriegeleitete Verlaufsanalyse

Die Bildung der theoriegeleiteten Verlaufgruppen wurde bereits in Abschnitt 4.3.4 beschrieben. Zur Überprüfung der ersten Hypothese wurden die Häufigkeitsverteilung sowie die Konfidenzintervalle der relativen Häufigkeiten der vier a priori definierten Verlaufgruppen berechnet.

Zum statistischen Vergleich der beiden Verlaufgruppenbildungen (explorativ und theoriegeleitet) kam der Übereinstimmungskoeffizient Cohens Kappa (Cohen, 1960) zum Einsatz. Nach Landis und Koch kann bei einem Kappa-Koeffizienten von 0.21 bis 0.40 von einer ausreichenden Übereinstimmung (fair agreement) gesprochen werden (Landis & Koch, 1977). Andere Autoren setzen konservativere Interpretationsgrenzen an und sprechen erst ab einem Kappa von 0.40 von einer ausreichend vorhandenen Übereinstimmung (Cicchetti, 1994; zitiert nach Wirtz & Kutschmann, 2007).

4.5.3 Prädiktoren- und Outcome-Analysen

Für die Untersuchung, ob zwischen den gefundenen Symptomverlaufsklassen Unterschiede in relevanten Patienten- und Prozessmerkmalen bestehen, wurden im Rahmen univariater Analysen Chi²-Tests für kategoriale Variablen und Varianzanalysen für metrische Variablen berechnet. Zur Überprüfung der prospektiven *prädiktiven Fragestellungen* kommen logistische und lineare Regressionen zur Anwendung.

Zur Untersuchung der Vorhersage der Gruppenzugehörigkeit zu den Symptomverlaufgruppen durch Patienten- und Prozessmerkmale (*Fragestellung 2*) wurden multinomiale logistische Regressionen berechnet. Diese Methode zielt auf die Ableitung einer Eintrittswahrscheinlichkeit für ein empirisch beobachtetes Ereignis (Backhaus, Erichson, Plinke & Weiber, 2003), welches hier durch den Symptomverlauf bzw. genauer die Verlaufgruppe abgebildet wird. Vorhergesagt wird demnach die Wahrscheinlichkeit der Zugehörigkeit zu einer bestimmten Symptomverlaufgruppe. Im Gegensatz zur linearen Regression ist der Zusammenhang zwischen den Prädiktoren und der kategorialen Variable (Kriterium) nicht linear sondern logistisch. Für jeden untersuchten Prädiktor werden neben den aus der linearen Regression bekannten β -Gewichten sogenannte Odds-Ratio-Werte berechnet. Dieser Wert gibt an, wie sich die Chance für die Gruppenzugehörigkeit ändert, wenn die unabhängige Variable bzw. der Prädiktor um eine Einheit erhöht wird. Mit Zu- bzw. Abnahme der Einheit im Prädiktor x (z.B. Alter mit der Einheit Jahre), steigt bzw. sinkt die Wahrscheinlichkeit für die Zuordnung zu der untersuchten Gruppe. Es wurden sowohl dichotome als auch intervallskalierte Variablen als Prädiktoren in das Modell aufgenommen.

Kategorial-mehrstufige Prädiktorvariablen (z.B. höchster Schulabschluss) wurden dichotomisiert.

Zum besseren Verständnis werden die Begriffe *Risiko*, *Chance* und *Odds Ratio* an einem Beispiel der Response auf Schmerzmittel- vs. Placebo-Therapie veranschaulicht. Dabei soll die Wahrscheinlichkeit der Zugehörigkeit zur Response-Gruppe (Therapieresponse vs. Non-Response als abhängige Variable) in Abhängigkeit von der erhaltenen Therapie (Schmerzmittel vs. Placebo als unabhängige Variable) vorhergesagt werden. Die Wahrscheinlichkeit für eine Response unter der Bedingung Schmerzmitteltherapie wird aus dem Anteil der Responder an der Gesamtanzahl der mit Schmerzmitteln behandelten Personen berechnet und wird auch als *Risiko*¹⁴ bezeichnet. Das *Relative Risiko* bezeichnet das Verhältnis der unter Schmerzmitteltherapie respondierenden zu den unter Placebo Therapie respondierenden Patienten. Ebenso kann man das Risiko für das Nicht-Respondieren unter beiden Therapiebedingungen berechnen. Setzt man das Risiko, unter Schmerztherapie zu respondieren ins Verhältnis zu dem Risiko, unter Schmerztherapie *nicht* zu respondieren, erhält man die sog. *Chance* einer Response unter Schmerztherapie. Ebenso wird die Chance einer Response unter Placebobedingungen berechnet. Die Chance wird auch als *Odds* bezeichnet. Setzt man die Chance für ein Ansprechen unter Schmerzmitteltherapie ins Verhältnis zu der Chance eines Ansprechens unter Placebothherapie, erhält man den Wert des Chancenverhältnisses, das *Odds Ratio*. Liegt dieser Wert bei 1, ist die Chance in beiden Gruppen (Ansprechen oder nicht Ansprechen) gleich groß, man spricht auch von einem Null-Effekt. Odds-Ratio-Werte über 1 sprechen für eine erhöhte Chance, unter der untersuchten Bedingung der Gruppe zuzugehören, Werte unterhalb von 1 zeigen eine verringerte Chance an, unter der untersuchten Bedingung der Gruppe zuzugehören. Die Parameter werden an einem Beispiel (Edwards, Moore & McQuay, 2004; zitiert nach Schwarzer, Türp & Antes, 2004)¹⁵ in der folgenden Vierfeldertafel (Tabelle 2) veranschaulicht.

¹⁴ Die Begriffe *Risiko* und *Chance* werden hier immer anhand der beschriebenen Definition verwendet, unabhängig davon, ob das Zielereignis positiv oder negativ ist.

¹⁵ In der Studie von Edwards et al. (2004) wird die Wirkung des Schmerzmittelwirkstoffs Diclofenac (50mg) mit der Wirkung eines Placebopräparates bei postoperativen Zahnschmerzen verglichen. Das Outcome ist die Reduktion der Zahnschmerzen um mindestens 50% 6 Stunden nach der Operation.

Tabelle 2: Vierfeldertafel zum Erfolg und Misserfolg zweier Therapiebedingungen

		Zielkriterium: Therapieerfolg		Gesamt
		Response	Non-Response	
Prädiktor: Therapiebedingung	Schmerzmittel	32	89	121
	Placebo	8	55	63
Gesamt		40	144	184

Beträgt das Risiko eines Ansprechens auf Schmerzmittel $32/121 = 0.26$ und das Risiko für ein Nichtansprechen entsprechend $89/121 = 0.74$, so liegt die Chance für ein Ansprechen unter Schmerzmitteltherapie bei $(32/121) / (89/121) = 32/89 = 0.36$. Dahingegen beträgt das Risiko für ein Ansprechen unter Placebothherapie nur $8/63 = 0.13$ und das Risiko für ein Nicht-Ansprechen 0.87 , sodass die Chance für ein Ansprechen unter Placebobedingungen bei 0.15 liegt. Das Odds Ratio für die Vorhersage der Gruppenzugehörigkeit *Response* durch den Prädiktor Schmerzmitteltherapie liegt demnach bei $0.36/0.15 = 2.4$. Das bedeutet, dass ein Patient unter Schmerzmitteltherapie eine 2.4fach erhöhte Chance für eine Response aufweist als ein Patient unter Placebothherapie.

Durch die Berechnung der Konfidenzintervalle der Odds Ratios kann die Genauigkeit der Schätzung berücksichtigt werden. Neben den Odds-Ratio-Werten der einzelnen Prädiktoren werden als Parameter der Modellgüte Pseudo-*R*-Quadrat-Statistiken, entsprechend dem Bestimmungsmaß R^2 der linearen Regressionsanalyse, berechnet. Dabei wird in der vorliegenden Arbeit das *Nagelkerke-R²* als Gütemaß herangezogen.

Zur Analyse des Zusammenhangs zwischen Symptomverlauf und Outcome (*Fragestellung 3*) kamen zum einen Kovarianzanalysen zum Einsatz. Dabei wurden die Unterschiede im Outcome zwischen den Verlaufgruppen unter statistischer Kontrolle des Ausgangswerts im jeweiligen Outcome-Maß (Prä-Wert des Outcomes) als Kovariate auf statistische Signifikanz getestet. Um den Zusammenhang von weiteren Confounder mit dem Outcome zu berücksichtigen, wurde die Fragestellung 3 zum anderen mittels linearer Regressionsanalysen untersucht. Dabei wurde überprüft, inwiefern die Outcome-Maße (primäres und sekundäres Outcome, beide als kontinuierliche abhängige Variablen) durch die Gruppenzugehörigkeit vorhergesagt werden können. Zusätzlich zu den im Fokus stehenden Symptomverläufen wurden weitere, theoretisch fundierte Prädiktoren des Outcomes in die Regressionsanalyse mit aufgenommen. Um den inkrementellen Beitrag zur Varianzaufklärung im Outcome durch die im Fokus stehenden Prädiktoren zu ermitteln, wurde eine hierarchische Vorgehensweise bei der Regressionsanalyse gewählt, bei der nicht

alle Prädiktoren gleichzeitig betrachtet, sondern inhaltlich begründete Prädiktorensätze in einer hierarchischen Folge nacheinander getestet werden. Diese Verfahren ermöglicht eine differenzierte Analyse des Zuwachses an Erklärungskraft durch einen bestimmten Prädiktor oder ein Prädiktorensatz, in Relation zum erklärten Varianzanteil des vorherigen Prädiktorensatzes.

Auf die Voraussetzungen der jeweiligen statistischen Verfahren und die Erfüllung dieser wird aufgrund der besseren Übersichtlichkeit vertiefend im Ergebnisteil eingegangen.

4.5.4 Bestimmung und Interpretation statistischer und klinischer Effekte

Zur Bestimmung des Signifikanzniveaus wurden gemäß den konventionellen Vereinbarungen Irrtumswahrscheinlichkeiten von $\alpha \leq .001$ als hoch signifikant, $\alpha \leq .01$ als sehr signifikant, $\alpha \leq .05$ als signifikant und $\alpha \leq .10$ als tendenziell signifikant eingestuft (Kennzeichnung wie folgt: $\alpha \leq .001^{***}$, $\alpha \leq .01^{**}$, $\alpha \leq .05^*$, $\alpha \leq .10^{\circ}$). Die tendenzielle Signifikanz auf dem 10%-Niveau wurde aufgrund des explorativen Charakters der ersten Fragestellung und den daraus resultierenden Fragestellungen 2 und 3 angewendet. Für Repräsentativitäts- und Drop-Out-Analysen wurde eine konservativere Vorgehensweise gewählt und tendenziell signifikante Ergebnisse unberücksichtigt gelassen. Auf eine Alpha-Adjustierung wurde trotz multiplen Testungen verzichtet: Aufgrund des explorativen Charakters der Untersuchungsfragestellungen wird eine möglichst hohe Teststärke angestrebt (vgl. Abschnitt 4.6), das Risiko einer Alpha-Fehler-Inflation wird zugunsten der Identifikation auch kleiner Effektstärken in Kauf genommen. Zur standardisierten Einordnung der Größe der erreichten Effektstärken werden die in Tabelle 3 aufgeführten jeweiligen Effektstärkenmaße angegeben. Für die Interpretation der Effektstärken werden die Konventionen nach Cohen (1988) zugrunde gelegt. Unterschieden werden kleine, mittlere und große Effektstärken (Kennzeichnung am Beispiel Cohens d : $d \geq .20^+$; $d \geq .50^{++}$; $d \geq .80^{+++}$). Die jeweiligen Grenzen der Effektmaße sind ebenfalls Tabelle 3 zu entnehmen.

Tabelle 3: Verwendete Effektstärkemaße und ihre Interpretation

Statistisches Verfahren	Effektstärkemaß	klein	mittel	groß
Mittelwertsunterschiede bei kontinuierlichen Variablen (<i>t</i> -Test)	Cohens <i>d</i>	≥ .20	≥ .50	≥ .80
Prä-Post-Vergleiche	Cohens <i>d</i> (SES)	≥ .20	≥ .50	≥ .80
Varianzaufklärung (Ko-/ Varianzanalyse)	partiell η^2	≥ .010	≥ .059	≥ .138
Verteilungsunterschiede bei kategorialen Variablen (Chi ² -Test)	Φ	≥ .10	≥ .30	≥ .50
Zusammenhang kontinuierlicher Variablen (Korrelation)	Pearsons <i>r</i>	≥ .10	≥ .30	≥ .50
Modellgüte multipler (hierarchischer) Regression	Cohens <i>f</i> ²	≥ .02	≥ .15	≥ .35
Modellgüte logistischer Regressionen ¹⁶	Pseudo- <i>R</i> ² (Nagelkerke)	≥ .20	≥ .40	≥ .50

Anmerkung: Die Beurteilung der Modellgüte des Nagelkerke-*R*² wird mit ≥ .20 als akzeptabel, ≥ .40 als gut und ≥ .50 als sehr gut bezeichnet.

Zusätzlich zu den oben beschriebenen Kovarianzanalysen werden Effektstärken der Veränderung zwischen Aufnahme und Katamnese berechnet. Für die Berechnung von Eingruppen-Prä-Post-Vergleichen wird die Mittelwertsdifferenz zwischen den untersuchten Messzeitpunkten an einem Streuungsmaß standardisiert. Bezüglich des Streuungsmaßes, das dabei angewendet wird, bestehen unterschiedliche Möglichkeiten und Empfehlungen, dies muss jeweils anhand der Fragestellung und der angestrebten Vergleiche entschieden werden (für einen Überblick vgl. z.B. Maier-Riehle & Zwingmann, 2000). Angewendet wird häufig die *Standardized Effect Size* (SES; Kazis, Anderson & Meenan, 1989), die die Standardabweichung der Prä-Werte einbezieht (z.B. Stiles et al., 2008a), die gepoolte Standardabweichung der Prä-Post-Werte (z.B. Hahlweg, Fiegenbaum, Frank, Schroeder & von Witzleben, 2001; Hartmann, Herzog & Drinkmann, 1992) oder die Standardabweichung der Differenzwerte ("standardizes response mean" SRM; Cohen, 1988).

In der vorliegenden Untersuchung wird zur Berechnung der Effektstärke die Streuung der Prä-Werte der Gesamtstichprobe gewählt. Die Prä-Streuung der (noch) unbehandelten Patienten scheint für das hier gewählte Design der beste Schätzer einer (hypothetischen) vergleichbaren unbehandelten Kontrollgruppe zu sein, da davon ausgegangen wird, dass die Streuung sich nicht verändern würde, wenn die Patienten keine Behandlung erhalten würden (im Sinne einer hypothetischen Kontrollgruppe). Zudem wird diese Vorgehensweise vor allem für heterogene Stichproben mit mehreren Indikationsgruppen empfohlen (Steffanowski et al., 2007).

¹⁶ Die Beurteilung der Modellgüte erfolgte in Anlehnung an Backhaus, Erichson, Plinke und Weiber, 2003, S. 447 f. (Backhaus et al., 2003).

Zur Bestimmung von Therapieeffekten wurde zudem das Konzept der klinischen Signifikanz angewendet (Jacobson & Truax, 1991). Demnach müssen zwei Kriterien der Veränderung erfüllt sein: Zum einen muss die Veränderung auf individueller Ebene die Mindestgröße einer *reliable change* erreichen (Reliable Change Index, RCI¹⁷). Diese stellt eine statistisch bedeutsame Messwertdifferenz unter Berücksichtigung der Messgenauigkeit des verwendeten Messinstrumentes dar. Ist diese kritische Differenz überschritten, spricht man von einer statistisch signifikanten Veränderung. Das zweite Kriterium stellt die Veränderung aus dem dysfunktionalen Wertebereich in den funktionalen Wertebereich (Überschreiten des sogenannten Cut-Off-Wertes) dar. Patienten, die sich in einem Mindestausmaß (RCI) vom klinisch auffälligen in den klinisch unauffälligen Bereich verändern, gelten als klinisch signifikant verbessert. Die Berechnung der klinischen Signifikanz birgt neben der Vergleichbarkeit mit anderen Studienergebnissen (Angabe von Häufigkeiten) die Vorteile, dass bei diesem Kennwert eine klinische Relevanz angezeigt ist, und, da der Kennwert für jeden Patienten berechnet wird, Aussagen über Veränderungen auf Einzelfallebene ermöglicht werden. Der Nachteil dieses Verfahrens ist, dass nicht mehr eine kontinuierliche Auswertung der Daten erfolgt, sondern anhand von Cut-Off-Werten Gruppen gebildet werden, womit eine Informationsreduktion in Kauf genommen wird. Aus diesem Grund wird in der vorliegenden Arbeit das Konzept der klinischen Signifikanz nur in Ergänzung zu kontinuierlichen Effektstärkenmaßen eingesetzt. Darüber hinaus impliziert das Konzept die Annahme, dass eine vollständige Heilung möglich ist und dies auch den Regelfall darstellt: Dies trifft aber für viele chronische Erkrankungen nicht zu, kleinere Verbesserungen sind hier häufig das primäre Therapieziel.

Für die Skalen des HEALTH-49 wurden RCI- und Cut-Off-Werte anhand von klinischen (stationäre Psychotherapiepatienten, vgl. auch Abschnitt 4.3.1) und gesunden Referenzstichproben von Rabung et al. (2008) berechnet und stehen zur Bestimmung der klinischen Signifikanz zur Verfügung.

Ein weiteres Kriterium, das zur Bestimmung von Symptomveränderung zum Einsatz kommt, ist die Reduktion der Symptombelastung um 50%. Dieses Kriterium wird aufgrund des in der Literatur häufig angewendeten prozentualen Verbesserungsausmaßes zusätzlich für die wöchentlichen Veränderungen bestimmt. Dabei wird der Wert der Symptombelastung zu einem kritischen Zeitpunkt (Woche x) durch den Wert der Symptombelastung zu Behandlungsbeginn geteilt und mit 100 multipliziert. Anschließend wird die Variable dichotomisiert, in dem jeder Patient entweder der Gruppe *Verbesserung um mindestens 50% erreicht* oder der Gruppe *Verbesserung um mindestens 50% nicht erreicht* zugeordnet wird. Der Informationsverlust, der mit der Dichotomisierung einhergeht wird zum Zweck der

¹⁷ Die Formel für die kritische Differenz (*Reliable Change Index; RCI*) lautet:
 $D_{\text{krit}} = z_{\alpha}(1.96) * SD_{\text{diff}} * \sqrt{2*(1-r)}$.

Vergleichbarkeit mit anderen Studien in Kauf genommen. Ferner dient dieses Kriterium lediglich der zusätzlichen Darstellung der Veränderungsverläufe und ist nicht ausschlaggebend für die Hauptfragestellungen.

4.6 Poweranalysen

4.6.1 Statistische Poweranalyse

Zur Ermittlung der benötigten Stichprobengröße, um tatsächlich bestehende Unterschiede bzw. Zusammenhänge mit einer definierten Wahrscheinlichkeit aufdecken zu können, wurde eine Poweranalyse mit der Software G*Power (Version 3.1.2; Erdfelder, Faul & Buchner, 1996) durchgeführt. Dabei bezieht sich die Analyse auf die Hauptfragestellung, nämlich die Vorhersage des längerfristigen primären und sekundären Outcomes aus den Symptomverläufen. Das zugrunde liegende Maß ist der inkrementelle Anteil der aufgeklärten Varianz durch die Symptomverläufe an der Gesamtvarianz, die durch alle in dem Modell eingeschlossenen Variablen aufgeklärt wird. Auf der Basis der bisherigen empirischen Ergebnisse wurden die Hypothesen formuliert, dass der inkrementelle Varianzanteil signifikant, also bedeutend größer als Null ist. Ferner wird angenommen, dass der inkrementelle Varianzanteil die Höhe einer kleinen bis mittleren Effektstärke ($f^2 = 0.02 - 0.15$; s. Tabelle 3) erreicht. Da es sich um eine explorative Fragestellung handelt, soll die Teststärke (Power: $1-\beta$) möglichst hoch sein, um die Wahrscheinlichkeit, mindestens kleine Effekte zu entdecken, zu erhöhen. Die Ergebnisse der Poweranalyse für eine hierarchische multiple Regression mit angenommenen 4 primär zu testenden Prädiktoren (Symptomverlaufgruppen) und insgesamt 25 Prädiktorvariablen (Patienten- und Prozessmerkmale, die als Confounder berücksichtigt werden) sind in Abbildung 3 dargestellt¹⁸.

¹⁸ Die Analyse zur Ermittlung der Stichprobe bezieht sich auf das Verhältnis der Varianzaufklärung durch die 4 primär zu testenden Prädiktoren zu der Varianzaufklärung durch alle 25 eingeschlossenen Prädiktoren.

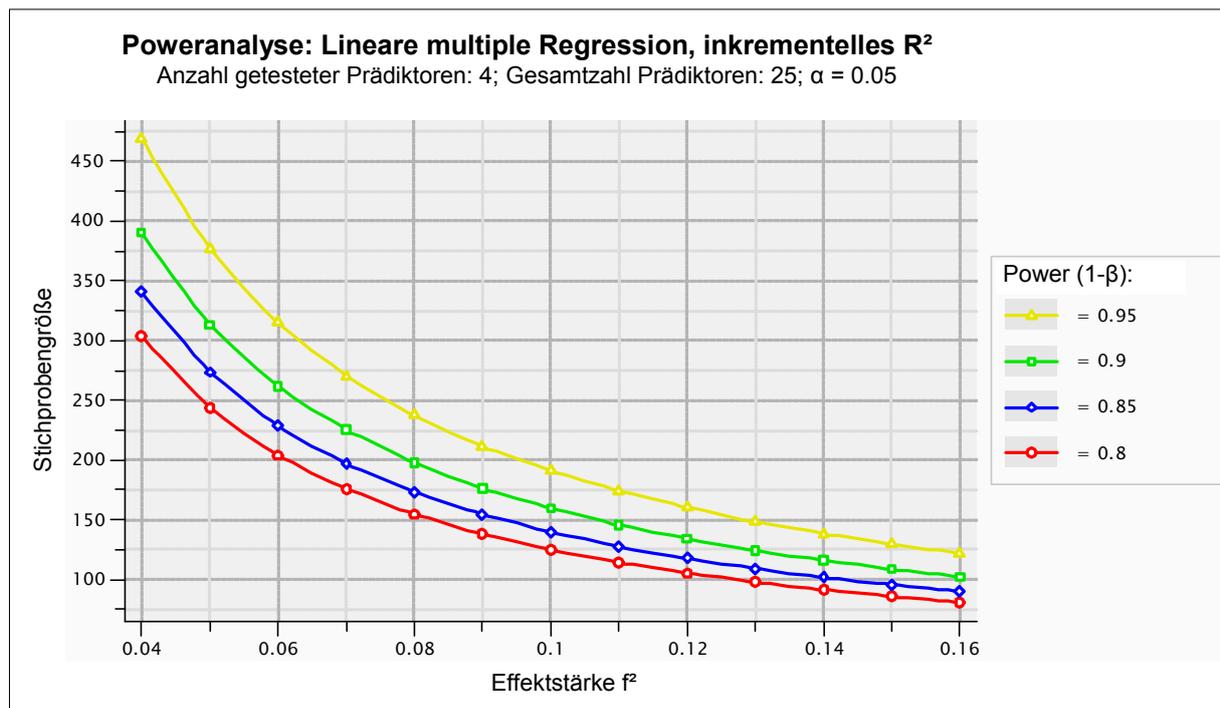


Abbildung 3: Statistische Poweranalyse für multiple hierarchische Regression

Um einen mindestens kleinen bis mittleren Effekt ($f^2 = 0.06$) aufzudecken, ist, bei einer Teststärke von mindestens 80% und einer Irrtumswahrscheinlichkeit α von $p < .05$, eine Stichprobengröße von $n = 205$ erforderlich. Da die grundlegende Fragestellung der vorliegenden Arbeit einen stark explorativen Charakter hat (Fragestellung 1), und die Analysen der Hauptfragestellung (Fragestellung 3) auf den Ergebnissen dieser beruhen, soll der Fehler, einen vorhandenen Effekt *nicht* zu entdecken, möglichst minimiert werden. Daher wird für die Stichprobenkalkulation eine höhere Power von mindestens 95% gewählt, womit sich die erforderliche Stichprobe auf $n = 315$ beläuft (s. Abbildung 3, gelbe Kurve). Mit einer Katamnese-Stichprobe von 315 Patienten können demnach kleine bis mittlere Effekte der inkrementellen Vorhersagekraft der Symptomverläufe auf das längerfristigen Behandlungsergebnis mit einer Fehlerwahrscheinlichkeit von $p < .05$ und einer Teststärke von mindestens 95% aufgezeigt werden.

4.6.2 Stichprobenkalkulation und Erhebungszeitraum

Ausgegangen wird von der Erhebung in drei stationären Fachkliniken. Zum Katamnesezeitpunkt sollen von mindestens 315 Patienten komplette, auswertbare Datensätze einschließlich der Verlaufsmessungen vorliegen. Als Einschlusskriterium für die Verlaufsanalyse müssen neben den Aufnahmedaten noch von zwei weiteren Messzeitpunkten im Verlauf oder am Ende der Behandlung Daten vorliegen (vgl. Abschnitt 4.5.2.1). Diese beiden Messzeitpunkte können zwei Verlaufsmessungen oder eine Verlaufs-

sowie die Entlassmessung sein. Für die Prozess-Outcome-Fragestellung muss zudem ein auswertbarer Datensatz zur Katamnese vorliegen. Um die für die Hauptfragestellung erforderliche Stichprobe von $n = 315$ zu erreichen, sind bei einer konservativ geschätzten Rücklaufquote von ca. 60% zur Katamnese in jeder Klinik ca. 180 Patienten einzubeziehen, die zur Aufnahme, während der Behandlung und zum Entlassungszeitpunkt untersucht werden. Bei ca. 10 Patienten pro Aufnahmewoche und einer durchschnittlichen Behandlungsdauer von 39 Tagen wurde von einem Mindesterhebungszeitraum von ca. 24 Wochen bzw. 6 Monaten pro Klinik ausgegangen.

5 Ergebnisse

Im Folgenden werden die Ergebnisse der drei Fragestellungen und der jeweiligen Hypothesen der vorliegenden Studie dargestellt.

Im ersten Abschnitt dieses Kapitels (5.1) erfolgt zunächst die Bildung und Beschreibung der Analyse-Stichproben. Anschließend werden die Untersuchungen zur Repräsentativität der Stichproben und zu potentiellen Selektionseffekten durch Patienten, die die Studienteilnahme im Laufe der Untersuchung abgebrochen haben (Drop-Out), beschrieben. In den Abschnitten 5.2 bis 5.5 werden die Ergebnisse der vier Unterfragestellungen der *ersten Fragestellung* (1 A bis 1 D) und der entsprechenden *Hypothese 1* dargestellt. Der Abschnitt 5.6 beinhaltet die Ergebnisse zur *zweiten Fragestellung* und den hier postulierten *Hypothesen 2* und *3*. In den fünf folgenden Abschnitten des Ergebniskapitels (5.7 – 5.11) werden die Ergebnisse zur *dritten Fragestellung* (Hauptfragestellung; 3 A – 3 D) und den jeweils formulierten *Hypothesen 4* bis *11* dargestellt.

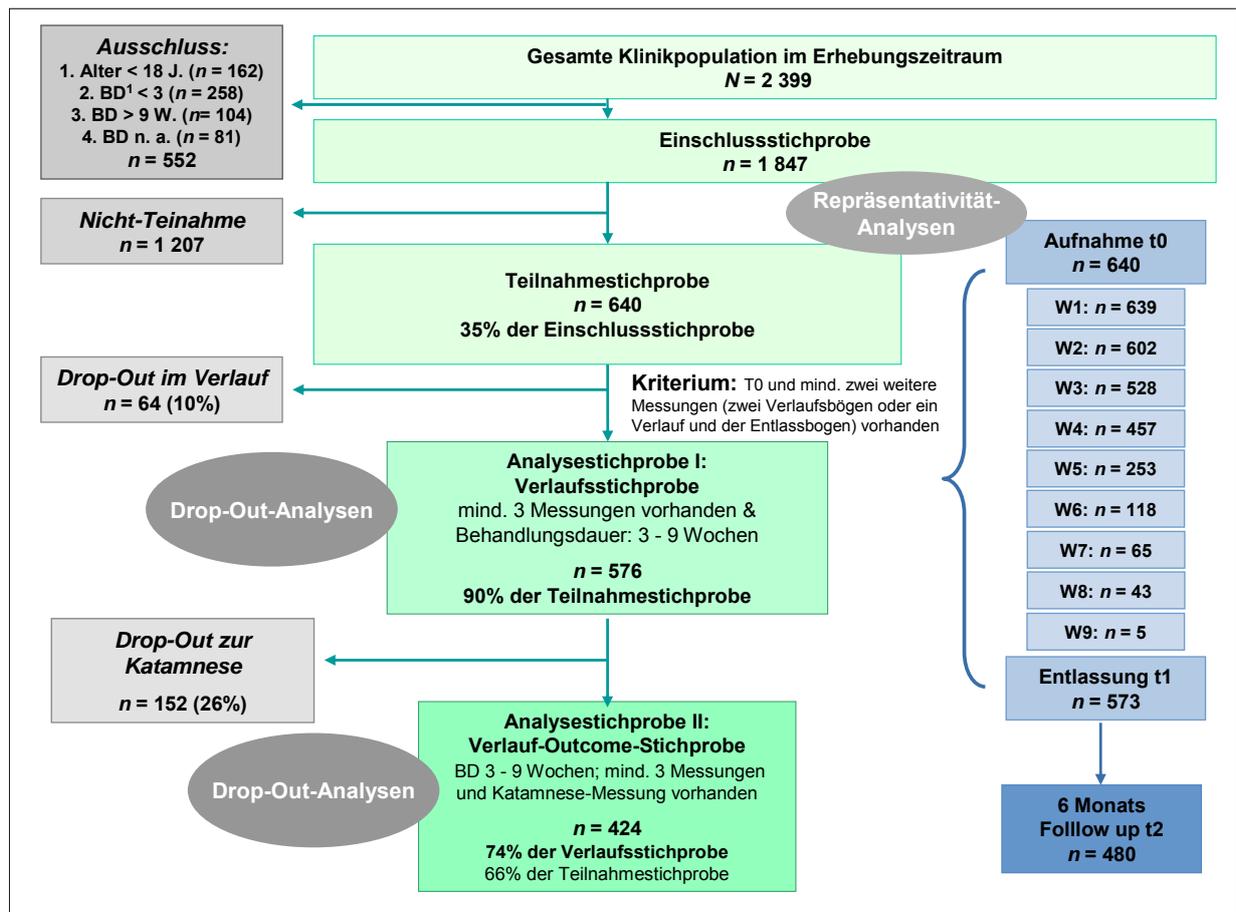
5.1 Stichproben, Repräsentativitäts- und Drop-Out-Analysen

5.1.1 Stichprobenbildung

Die Datenerhebung begann im Mai 2008 und erfolgte sukzessiv in drei Kliniken. Der durchschnittliche Erhebungszeitraum betrug 10 Monate, insgesamt wurden Daten über den Zeitraum von 23 Monaten erhoben. Der Erhebungszeitraum endete im März 2010. Im gesamten Erhebungszeitraum (ausschließlich der Katamnese-Erhebung) wurden in den drei Kliniken insgesamt 2 399 Patienten behandelt. Nach Prüfung der Ein- und Ausschlusskriterien wurde die Grundgesamtheit um fast ein Viertel der Patienten ($n = 552$; 23%) reduziert. Einen großen Anteil nehmen hierbei Patienten mit einer Behandlungsdauer von weniger als 3 Wochen ($n = 258$) ein, wobei hier auch irregulär abgebrochene Behandlungen enthalten sind. Weitere 104 Patienten wurden länger als 9 Wochen behandelt. Von 81 Patienten war die Behandlungsdauer nicht bekannt¹⁹. Die Anzahl der Patienten, die das Erwachsenenalter noch nicht erreicht hatten, belief sich auf 162. Entsprechend stand eine Einschlussstichprobe von 1 847 Patienten zur Verfügung. Aus den drei Kliniken konnte eine konsekutive Stichprobe von insgesamt 640 Patienten für die Studienteilnahme gewonnen werden (35% der Einschlussstichprobe). Während der

¹⁹ Die Patientendaten der Klinikpopulation wurden von der jeweiligen Klinik anonymisiert übermittelt. Von einem Teil der Patienten ($n = 81$), die nicht an der Studie teilgenommen haben, konnte die Behandlungsdauer nur im Durchschnitt, nicht aber personenbezogen übermittelt werden. Da die Behandlungsdauer von den Patienten, die an der Studie teilgenommen haben, für die Studie dokumentiert wurde, handelt es sich bei diesen 81 Patienten ausschließlich um Nichtteilnehmer.

Erhebung betrug der Drop-Out 10%, von 64 Patienten waren weniger als drei Messzeitpunkte vorhanden. Die Verlaufsstichprobe umfasst somit 576 Patienten. Alle Patienten der Verlaufsstichprobe wurden 6 Monate nach Beendigung der Behandlung zur Katamnese-Erhebung angeschrieben. Mit einem Rücklauf von 74% steht eine Verlaufs-Outcome-Stichprobe von 424 Patienten mit vollständigen Verlaufs- und Katamnese-Daten zur Verfügung. Die Bildung der beiden Analysestichproben ist anhand eines Patientenflussdiagramms in Abbildung 4 dargestellt.



¹ Behandlungsdauer

Abbildung 4: Patientenflussdiagramm, Repräsentativitäts- und Drop-Out-Analysen

5.1.2 Repräsentativitätsanalysen

Im gesamten Erhebungszeitraum wurden 1 847 Patienten in den Kliniken behandelt, die die Einschlusskriterien erfüllten. Von den potentiellen Studienteilnehmern hat rund ein Drittel der Patienten einer Studienteilnahme zugestimmt (n = 640). Um zu prüfen, wie repräsentativ die Teilnehmer der Studie für die Klinikpopulation sind, wurden diese mit den Nicht-Teilnehmern (n = 1 207) aus der Einschlussstichprobe hinsichtlich der Merkmale Alter, Geschlecht, Behandlungsdauer und erster Entlassdiagnose verglichen. Die Ergebnisse der Repräsentativitätsanalysen sind in Tabelle 4 dargestellt.

Tabelle 4: Vergleich der Teilnehmer und der Nicht-Teilnehmer

Merkmal	Teilnehmer (n = 640)	Nicht-Teilnehmer (n = 1 207)	Signifikanztest
	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test
Anteil Frauen	76.8	75.8	$\chi^2 (1, n = 1 845) = 0.227$ $p = .633, \phi = -.011$
Alter; M (SD)	43.7 (11.3)	43.3 (12.0)	$t(1 845) = -0.765$ $p = .445, d = 0.03$
Behandlungsdauer in Tagen; M (SD)	41.8 (8.8)	40.2 (9.6)	$t(1 845) = -3.687$ $p < .001^{***}, d = 0.16$
Erste Entlassdiagnose (ICD-10 Kodierung)	%	%	
Unipolare Depressionen (F32, F33, F34.1)	55.9	68.7	
Anpassungsstörungen (F43.2)	16.9	7.1	
Angststörungen (F40, F41, F48)	7.2	5.0	$\chi^2 (1, n = 1 813) = 59.855$ $p < .001^{***}, \phi = .182^+$
Somatoforme Störungen (F45)	3.7	2.6	
Essstörungen (F50)	7.3	8.9	
Andere psych. Störungen	4.9	4.5	
Adipositas (E66)	4.1	3.2	

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**), und große (***) Effektstärken interpretiert.

Für Alter und Geschlecht zeigen sich keine signifikanten Unterschiede zwischen Teilnehmern und Nicht-Teilnehmern. Die Behandlungsdauer liegt bei den teilnehmenden Patienten etwas höher als bei den Nicht-Teilnehmern. Dieser Unterschied erweist sich zwar als signifikant, die Effektstärke erreicht jedoch nicht das Ausmaß einer kleinen Effektstärke ($d = 0.16$). Die Verteilung der Erstdiagnosen unterscheidet sich signifikant zwischen den Gruppen mit dem Ausmaß einer kleinen Effektstärke. Bei Betrachtung der einzelnen Diagnosegruppen zeigt sich, dass dieser Unterschied vor allem durch die Indikationen Depression und Anpassungsstörung bedingt ist. Im Vergleich zu den Nicht-Teilnehmern sind in der Gruppe der Teilnehmer weniger depressive Patienten und mehr Patienten mit Anpassungsstörung. In beiden Stichproben stellen die beiden Indikationen mit 73% und 76% den größten Anteil an Erstdiagnosen dar. Alle anderen Diagnosegruppen unterscheiden sich nicht wesentlich zwischen den Gruppen. Die Verteilung der Diagnosen in den beiden Stichproben ist in Abbildung 5 veranschaulicht.

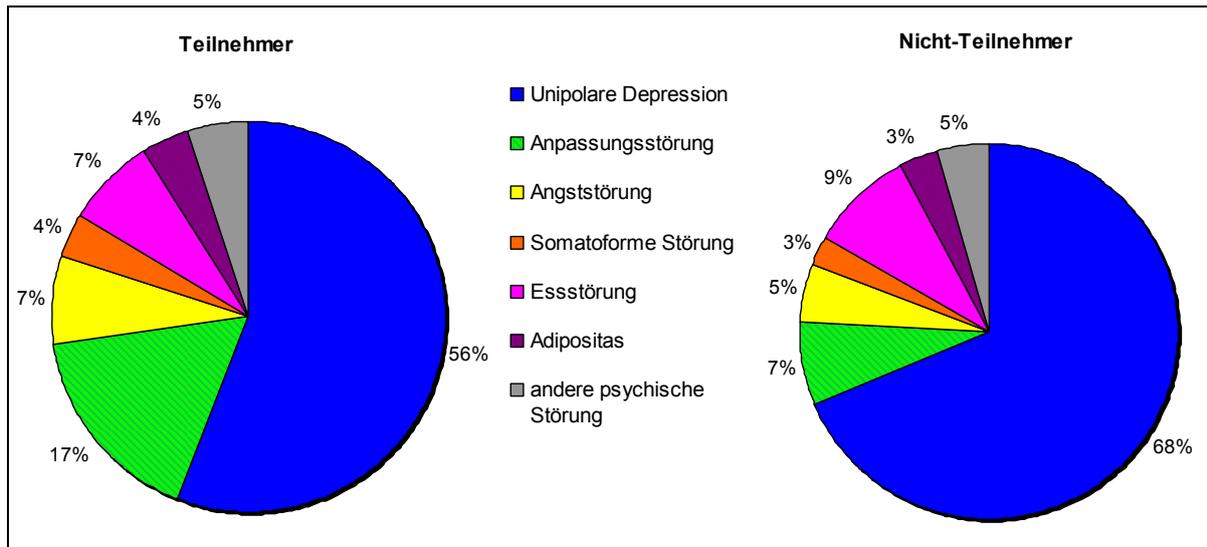


Abbildung 5: Verteilung der Erstdiagnosen der Teilnehmer und der Nicht-Teilnehmer

5.1.3 Stichprobenbeschreibung und Drop-Out-Analysen

5.1.3.1 Verlaufsstichprobe und Drop-Out im Behandlungsverlauf

Zur Aufnahme haben sich insgesamt 640 Patienten zur Teilnahme an der Studie bereit erklärt und einen Aufnahme-Bogen ausgefüllt. Alle Patienten, die im weiteren Verlauf mindestens zwei Verlaufsfragebögen oder mindestens einen Verlaufsfragebogen sowie den Entlassfragebogen ausgefüllt haben bilden die Verlaufsstichprobe. Unter diesen Bedingungen konnte eine Verlaufsstichprobe von $n = 576$ Patienten eingeschlossen werden. Die Merkmale der Verlaufsstichprobe, die zur Untersuchung der ersten beiden Fragestellungen (Prozessanalysen) herangezogen wird, sind in Tabelle 5 bis Tabelle 8 dargestellt. Um mögliche Selektionseffekte durch die Studienabbrecher zu untersuchen, wurden statistische Unterschiedstestungen zwischen der Verlaufsstichprobe und den Studienabbrechern ($n = 64$; s. Abbildung 4) durchgeführt, obwohl der mit 10% geringe Anteil der Studienabbrecher keine wesentlichen Selektionseffekte erwarten lässt. In Tabelle 5 werden die beiden Stichproben hinsichtlich soziodemographischer Merkmale, in Tabelle 6 hinsichtlich klinischer und sozialmedizinischer Merkmale und in Tabelle 7 hinsichtlich der Diagnosen vergleichend dargestellt.

Die Verlaufsstichprobe zeichnet sich durch einen für das Setting typischen hohen Anteil an Frauen und einem durchschnittlichen Alter von 55 Jahren aus. Knapp die Hälfte der Patientenstichprobe ist verheiratet (45%) und weitere 13% leben in einer festen Partnerschaft. Das Bildungsniveau liegt im mittleren Bereich, 24% der Patienten erreichten als höchsten Schulabschluss die Hauptschule, 41% den Realschulabschluss und 32% das Abitur oder Fachabitur. Ein ähnliches Bild zeigt sich hinsichtlich des höchsten Berufsabschlusses, bei dem der größte Anteil durch eine abgeschlossene Lehre

repräsentiert ist. Knapp ein Viertel der Patienten ist zum Zeitpunkt der Aufnahme nicht erwerbstätig. Entsprechend verteilt sich auch das zur Verfügung stehende monatliche Haushaltsnettoeinkommen, 14% der Patienten stehen nur bis zu 1 000 Euro im Monat zur Verfügung, 33% verfügen über bis zu 2 000 Euro, 25% über bis zu 3 000 Euro.

Die Drop-Out-Analysen ergaben in allen soziodemographischen Merkmalen keine signifikanten Unterschiede bzw. keine Unterschiede mit mindestens kleiner Effektstärke zwischen der Verlaufsstichprobe und denjenigen Patienten, die im Verlauf die Teilnahme an der Studie abbrechen. Demzufolge liefern die Analysen keinen Hinweis auf Selektivität der Verlaufsstichprobe (s. Tabelle 5).

Tabelle 5: Soziodemographische Patientenmerkmale der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
Soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Geschlecht				
Frauen	490 (76.8)	447 (77.6)	43 (69.4)	$\chi^2 (1, n = 638) = 2.138$ $p = .144, \phi = -.058$
Männer	148 (23.2)	129 (22.4)	19 (30.6)	
Alter				
in Jahren (Range: 18 – 77)	43.7 (11.3)	43.9 (11.1)	41.8 (12.9)	$t(635) = -1.382$ $p = .168, d = 0.18$
Familienstand/Partnerschaft				
mit Partner (verheiratet)	227 (44.5)	254 (45.0)	23 (39.7)	$\chi^2 (3, n = 622) = 1.579$ $p = .664, \phi = .050$
mit Partner (ledig)	82 (13.2)	72 (12.8)	10 (17.2)	
ohne Partner (ledig)	89 (14.3)	82 (14.5)	7 (12.1)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	174 (28.0)	156 (27.7)	18 (31.0)	
Schulabschluss				
Hauptschule	148 (23.7)	133 (23.7)	15 (23.8)	$\chi^2 (3, n = 625) = 2.046$ $p = .563, \phi = .057$
Realschule/ Polytechnische Oberschule	262 (41.9)	231 (41.1)	31 (49.2)	
Abitur/Fachabitur	193 (30.9)	178 (31.7)	15 (23.8)	
sonstige oder ohne Schulabschluss	22 (3.5)	20 (3.6)	2 (3.2)	
Berufsabschluss				
noch in Ausbildung	13 (2.2)	11 (2.1)	2 (3.6)	$FE (n = 586) = 5.972$ $p = .309, \phi = .101$
Lehre/Fachschule	328 (56.0)	299 (56.3)	29 (52.7)	
Meister/Techniker	81 (13.8)	70 (13.2)	11 (20.0)	

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
Soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	t-Test / χ^2-Test / FE-Test
Berufsabschluss (Fortsetzung)				
Fachhochschule/ Universität	101 (17.2)	96 (18.1)	5 (9.1)	
ohne Berufsabschluss	50 (8.5)	43 (8.1)	7 (12.7)	
sonstiges	13 (2.2)	12 (2.3)	1 (1.8)	
Erwerbstätigkeit				
Vollzeit berufstätig	302 (52.4)	274 (52.6)	28 (50.9)	
Teilzeit berufstätig	142 (24.7)	128 (24.6)	15 (25.5)	$\chi^2 (2, n = 576) = 0.056$ $p = .972, \phi = .010$
nicht berufstätig	132 (22.9)	119 (22.8)	13 (23.6)	
monatliches Haushaltsnettoeinkommen				
bis 1 000 Euro	95 (14.8)	81 (14.1)	14 (21.9)	
bis 2 000 Euro	210 (32.8)	189 (32.8)	21 (32.8)	
bis 3 000 Euro	160 (25.0)	143 (24.8)	17 (26.6)	$\chi^2 (4, n = 640) = 5.549$ $p = .235, \phi = .093$
bis 4 000 Euro	75 (11.7)	72 (12.5)	3 (4.7)	
mehr als 4 000 Euro	100 (15.6)	91 (15.8)	9 (14.1)	

M: Mittelwert; *SD*: Standardabweichung; χ^2 : Chi²-Test; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (***) und große (****) Effektstärken interpretiert.

Die durchschnittliche Behandlungsdauer der Patienten der Verlaufsstichprobe beträgt 42 Tage bzw. 6 Wochen mit einer Standardabweichung von 9 Tagen. Auch unter Berücksichtigung der Einschränkung der Varianz der Behandlungsdauer, bedingt durch den Ausschluss aller Patienten mit weniger als 3 und mehr als 9 Behandlungswochen, entspricht die mittlere Dauer den für das Setting üblichen Behandlungszeiten (vgl. Abschnitt 2.1.2). Somit kann die angestrebte erhöhte Vergleichbarkeit mit dem „prototypischen“ Behandlungsumfang (hinsichtlich der Dauer) angenommen werden, ohne dass die Varianz in der Behandlungsdauer für weitere Analysen zu gering wäre (vgl. Abschnitt 4.2.3). In den letzten 6 Monaten vor der stationären Behandlung waren die Patienten der Verlaufsstichprobe laut Selbstangabe durchschnittlich knapp 10 Wochen arbeitsunfähig. Diese Frage bezog sich sowohl auf erwerbstätige als auch auf nicht erwerbstätige Patienten, die Ihre Krankheitszeiten bezüglich der Einschränkungen bei der Ausübung der täglichen Aufgaben einschätzen sollten. Etwa 50% der untersuchten Patienten hat bereits psychotherapeutische Vorerfahrung durch ambulante Psychotherapie, wohingegen nur 10%

der Patienten aufgrund der derzeitigen psychischen Erkrankung schon einmal stationär behandelt wurden.

Für die Beschreibung der Stichprobe sowie die Untersuchung selektiver Effekte durch den Drop-Out stehen zudem psychometrische Skalen der Symptombelastung, gesundheitsbezogenen Lebensqualität sowie der allgemeinen Selbstwirksamkeitserwartung zur Verfügung. Für alle drei Skalen liegen Daten von Referenzgruppen oder bevölkerungsrepräsentativen Normstichproben vor. Mittels Ein-Stichproben-*t*-Test wurde für die jeweilige Skala geprüft, ob sich die Mittelwerte der Stichprobe signifikant von denen der Referenzgruppen unterscheiden.

Die durch die Patienten subjektiv eingeschätzte Ausprägung der Symptomatik zu Behandlungsbeginn liegt mit einem durchschnittlichen Skalenwert im HEALTH-49 Modul A (*Psychische und somatoforme Beschwerden*) von 1.41 im Mittel deutlich oberhalb des Cut-Off-Wertes einer klinisch bedeutsamen Symptombelastung (Cut-Off = 0.685) und unterscheidet sich nicht signifikant von dem Normmittelwert stationärer Psychotherapiepatienten bei Aufnahme [HEALTH-49 $A_{\text{stat. Psychotherapie}}$ ²⁰: $M = 1.45$ ($SD = 0.77$); $t(575) = -1.332$, $p = .183$, $d = 0.05$] (Rabung et al., 2008). Allerdings liegen zu Behandlungsbeginn fast 20% der Patienten unterhalb des den Cut-Off-Wertes der Symptombelastung. Vergleicht man die Stichprobenkennwerte jedoch mit Referenzdaten einer Stichprobe gesunder Probanden, erweisen sich die Unterschiede als hoch signifikant [HEALTH-49 A_{Gesunde} ²¹: $M = 0.34$ ($SD = 0.35$), $t(575) = 32.750$, $p < .001^{***}$, $d = -3.06$].

Die Skala *Psychischen Aspekte der gesundheitsbezogenen Lebensqualität* des SF-8 weist in der Untersuchungsstichprobe mit einem durchschnittlichen Skalenwert von 32.08 erwartungskonform ein deutlich niedrigeres Niveau als das einer gesunden Normstichprobe auf [SF-8 Normwerte ²²: $M = 53.25$ ($SD = 7.82$), $t(565) = 41.377$, $p < .001^{***}$, $d = 2.71$].

Die allgemeine Selbstwirksamkeitserwartung, erhoben mit dem SWE, liegt in der Verlaufsstichprobe mit einem Mittelwert von 23.48 ($SD = 6.51$) ebenfalls in einem deutlich niedrigeren Bereich als die Werte einer bevölkerungsrepräsentativen Normpopulation [SWE Normwerte ²³: $M = 29.43$ ($SD = 5.36$), $t(571) = -21.541$, $p < .001^{***}$, $d = 1.11$].

Sowohl hinsichtlich der Behandlungsdauer, der Arbeitsunfähigkeitszeiten und der psychotherapeutischen Vorbehandlung als auch in den Skalen der Symptombelastung, der gesundheitsbezogenen Lebensqualität und der allgemeinen Selbstwirksamkeitserwartung

²⁰ Die Referenzdaten basieren auf einer Stichprobe von $N = 1\,523$ Patienten aus 12 Kliniken der stationären psychosomatischen Rehabilitation (Rabung et al., 2008).

²¹ Die Referenzdaten basieren auf einer Stichprobe von $N = 2\,284$ gesunden Probanden aus 806 Hausarztpraxen der primärärztlichen Versorgung (Rabung et al., 2008).

²² Die Referenzdaten basieren auf einer repräsentativen Normstichprobe der deutschen Bevölkerung (Mehrthemenbefragung der Universität Leipzig; $N = 2\,511$) aus dem Jahr 2004 (Beierlein et al., 2010).

²³ Die Referenzdaten basieren auf einer Stichprobe von $N = 2\,019$ gesunden Probanden im Alter von 16 - 94 Jahren (Hinz et al., 2006).

bei Behandlungsbeginn treten keine signifikanten Unterschiede zwischen der Verlaufsstichprobe und der Drop-Out-Stichprobe auf. Entsprechend kann davon ausgegangen werden, dass auch in den krankheitsbezogenen Merkmalen keine selektiven Effekte durch den Abbruch der Studienteilnahme vorliegen (s. Tabelle 6).

Tabelle 6: Sozialmedizinische und klinische Patientenmerkmale der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
Sozialmedizinische/ klinische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	t-Test / χ^2-Test
Behandlungsdauer				
in Tagen (Range 19 – 60)	41.8 (8.8)	42.0 (8.5)	40.6 (10.9)	$t(638) = -1.155$ $p = .249, d = 0.16$
Arbeitsunfähigkeits-/ Krankheitszeiten (6 Monate vor t0)				
in Wochen (Range 0.2 - 26)	9.8 (9.4)	9.8 (9.4)	9.2 (9.5)	$t(591) = -.494$ $p = .621, d = 0.06$
Psychotherapeutische Vorbehandlung (6 Monate vor t0)				
Ambulante Vorbehandlung	300 (47.5)	277 (48.5)	23 (37.7)	$\chi^2 (1, n = 632) = 2.581$ $p = .108, \phi = .064$
keine ambulante Vorbehandlung	332 (52.5)	294 (51.5)	38 (62.3)	
Stationäre Vorbehandlung	67 (10.5)	58 (10.1)	9 (14.3)	$\chi^2 (1, n = 637) = 1.054$ $p = .304, \phi = -.041$
keine stationäre Vorbehandlung	570 (89.5)	516 (89.9)	54 (85.7)	
Symptombelastung				
HEALTH-49 A t0 (Range 0 – 3.72)	1.40 (0.78)	1.41 (0.78)	1.32 (0.79)	$t(634) = -0.810$ $p = .418, d = -0.12$
HEALTH-49 A t0: unterhalb des Cut-Off- Wertes (0.685)	125 (19.7%)	112 (19.4%)	13 (21.7%)	$\chi^2 (1, n = 636) = 0.170$ $p = .680, \phi = .016$
Gesundheitsbezogene Lebensqualität				
SF-8 Psych. t0 (Range 7.43 – 61.36)	32.13 (12.11)	32.08 (12.16)	32.60 (11.80)	$t(622) = 0.312$ $p = .755, d = -0.04$
Selbstwirksamkeitserwartung				
SWE t0 (Range 10 - 40)	23.59 (6.61)	23.48 (6.51)	24.61 (7.48)	$t(632) = 1.279$ $p = .201, d = -0.17$

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (***) und große (****) Effektstärken interpretiert.

Die Verteilung der Diagnosegruppen und Komorbidität in der Verlaufsstichprobe sowie in der Drop-Out-Stichprobe ist der Tabelle 7 zu entnehmen. Dabei werden die Diagnosegruppen aus allen vergebenen Entlassdiagnosen gebildet, sodass es in jeder Stichprobe zu einer aufsummierten Häufigkeit von über 100% kommen kann. Zusätzlich wird die Verteilung der ersten Entlassdiagnose dargestellt.

Über zwei Drittel aller Patienten weisen (allein oder komorbid) die Diagnose einer *unipolaren Depression* (ICD-10-Kategorien: F32, F33, F34.1) auf. Mit jeweils 14-18% sind außerdem die Diagnosegruppen *Anpassungsstörungen* (F43.2), *Angststörungen* (F40, F41, F48), *Essstörungen* (F50) und *Somatoforme Störungen* (F45) am stärksten vertreten. Weitere (Ko-)Diagnosen stellen die *Posttraumatische Belastungsstörungen* und *Persönlichkeitsstörungen* dar. Alle anderen F-Diagnosen kommen selten (< 5%) oder gar nicht vor, und wurden unter *Sonstige Psychische Störungen* subsumiert. Die Verteilung der Diagnosegruppen steht im Einklang mit dem Diagnosespektrum des untersuchten Versorgungsbereichs (vgl. Abschnitt 2.1.1).

Betrachtet man ausschließlich die erste vergebene Entlassdiagnose bilden die unipolaren Depressionen mit 57% ebenfalls die größte Gruppe, gefolgt von den Anpassungsstörungen (16%), den Angststörungen (7%) und den Essstörungen (7%) aus dem F-Kapitel des ICD-10. Die bei mehr als einem Viertel der Stichprobe vorkommende somatische (Ko-)Diagnose einer Adipositas (26%) war nur bei 22 Patienten (4%) die Erstdiagnose.

Bei mehr als der Hälfte der Patienten der Verlaufsstichprobe besteht eine psychische Komorbidität (mehr als eine F-Diagnose). Durchschnittlich wurden bei den Patienten 1.6 F-Diagnosen mit einer Standardabweichung von 0.9 vergeben.

In den Drop-Out-Analysen zeigten sich für zwei Indikationsgruppen signifikante Unterschiede zwischen Verlaufs- und Drop-Out-Stichprobe. Während in der Verlaufsstichprobe deutlich mehr unipolare Depressionen (67%) als bei den Studienabbrechern (49%) diagnostiziert wurden, fanden sich bei den Studienabbrechern deutlich mehr Anpassungsstörungen (32%) als in der Verlaufsstichprobe (18%). Nahezu das gleiche Bild zeigt sich bei der Betrachtung der ersten Entlassdiagnose. Die Unterschiede über alle Diagnosegruppen sind von kleiner Effektstärke aber nur tendenziell signifikant (Fishers exakter Test: $p = .069$; s. Tabelle 7), was jedoch durch die geringen Patientenzahlen der Drop-Out-Stichprobe erklärt werden kann.

Tabelle 7: Diagnosenverteilung der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe

Diagnose	ICD-10 Code	Gesamt (N = 618)	Verlaufs- Stichprobe (n = 571)	Drop-Out- Stichprobe (n=47)	Signifikanztest
		n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Alle vergebenen Entlassdiagnose (1. – 10.)					
Unipolare Depression	F32, F33, F34.1	408 (66.0%)	385 (67.4%)	23 (48.9%)	$\chi^2 (1, n = 618) = 6.617$ $p = .010^*$, $\phi = .103^+$
Angststörung	F40, F41, F48	113 (18.3%)	105 (18.4%)	8 (17.0%)	$\chi^2 (1, n = 618) = 0.054$ $p = .816$, $\phi = .009$
PTBS	F43.1	33 (5.3%)	31 (5.4%)	2 (4.3%)	$\chi^2 (1, n = 618) = 0.118$ $p = .713$, $\phi = .014$
Anpassungsstörung	F43.2	118 (19.1%)	103 (18.0%)	15 (31.9%)	$\chi^2 (1, n = 618) = 5.413$ $p = .020^*$, $\phi = -.094$
Somatoforme Störung	F45	87 (14.1%)	81 (14.2%)	6 (12.8%)	$\chi^2 (1, n = 618) = 0.072$ $p = .788$, $\phi = .011$
Essstörung	F50	102 (16.5%)	96 (16.8%)	6 (12.8%)	$\chi^2 (1, n = 618) = 0.516$ $p = .473$, $\phi = .029$
Persönlichkeitsstörung	F60 – F61	41 (6.6%)	36 (6.3%)	5 (10.6%)	FE (n = 618) = 1.317 $p = .228$, $\phi = -.046$
Andere psych. Störung		85 (13.8%)	79 (13.8)	6 (12.8%)	$\chi^2 (1, n = 618) = 0.042$ $p = .838$, $\phi = .008$
Adipositas	E66	164 (26.5)	151 (26.4)	13 (27.7)	$\chi^2 (1, n = 618) = 0.033$ $p = .856$, $\phi = -.007$
Erste Entlassdiagnose					
Unipolare Depression	F32, F33, F34.1	343 (55.9)	325 (57.2)	18 (39.1)	
Angststörung	F40, F41, F48	44 (7.2)	42 (7.4)	2 (4.3)	
PTBS	F43.1	11 (1.8)	11 (1.9)	0	
Anpassungsstörung	F43.2	104 (16.9)	90 (15.8)	14 (30.4)	
Somatoforme Störung	F45	23 (3.7)	21 (3.7)	2 (4.3)	FE (n = 614) = 12.885 $p = .069$, $\phi = .146$
Essstörung	F50	45 (7.3)	41 (7.2)	4 (8.7)	
Persönlichkeitsstörung	F60 - F61	9 (1.5)	7 (1.2)	2 (4.3)	
Andere psych. Störung		10 (1.6)	9 (1.6)	1 (2.2)	
Adipositas	E66	25 (4.1)	22 (3.9)	3 (3.6)	
Psychische Komorbidität					
Anzahl der F-Diagnosen (Range 0 - 5)		1.6 (0.9)	1.6 (0.9)	1.5 (0.8)	t(616) = -0.843 $p = .400$, $d = 0.11$
Mehr als eine F-Diagnose		343 (55.5)	317 (55.5)	26 (55.3)	$\chi^2 (1, n = 618) = 0.001$ $p = .979$, $\phi = .001$

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (***) und große (****) Effektstärken interpretiert.

Des Weiteren wurden die Drop-Out-Raten der drei Kliniken verglichen. Hier zeigen sich signifikanten Unterschiede zwischen der Verlaufsstichprobe und der Drop-Out-Stichprobe:

Während in Klinik 1 und 2 nur jeweils 8% der Patienten die Studienteilnahme abbrachen, lag die Drop-Out-Quote in Klinik 3 mit 16% doppelt so hoch. Die 64 Patienten der Drop-Out-Stichprobe gehen entsprechend zu unterschiedlich hohen Anteilen auf die Kliniken zurück. Die Drop-Out-Raten der drei Kliniken sind in Tabelle 8 dargestellt.

Tabelle 8: Vergleich der Drop-Out-Raten im Verlauf zwischen den drei Kliniken

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
Klinik	n (%)	n (%)	n (%)	χ^2-Test
Klinik 1	202 (100.0)	186 (92.1)	16 (7.9)	$\chi^2 (2, n = 640) = 8.392$ $p = .015^*$, $\phi = .115^+$
Klinik 2	257 (100.0)	237 (92.2)	20 (7.8)	
Klinik 3	181 (100.0)	153 (84.5)	28 (15.5)	

χ^2 : Chi²-Test; p : Signifikanz; ϕ : Effektstärke Phi; d : Effektstärke Cohens d . Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (****) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Insgesamt lassen sich in fast allen untersuchten Merkmalen keine bedeutsamen Unterschiede zwischen der Verlaufsstichprobe und der Drop-Out-Stichprobe finden. Lediglich in der Verteilung der beiden Indikationen Unipolare Depressionen und Anpassungsstörungen treten Selektionseffekte auf, die bei der Interpretation der Ergebnisse berücksichtigt werden müssen. Zudem differieren die Studienabbrucherquoten in den drei untersuchten Kliniken bedeutsam.

5.1.3.2 Verlaufs-Outcome-Stichprobe und Drop-Out zur Katamnese

Sechs Monate nach Ende der stationären Behandlung wurden die Patienten der Verlaufsstichprobe zur Katamnese-Erhebung angeschrieben. Von den 576 angeschriebenen Patienten antworteten 424 Patienten (74%), sodass von dieser Stichprobe Daten für die Verlaufs-Outcome-Analysen (Fragestellung 3) vorliegen. Um potentielle Selektionseffekte durch den Studienabbruch zur Katamnese zu berücksichtigen, werden für die 152 Patienten, die nicht zur Katamnese geantwortet haben (Katamnese-Drop-Out-Stichprobe), ebenfalls Drop-Out-Analysen für soziodemographische und krankheitsbezogene Merkmale durchgeführt. Die deskriptiven und inferenzstatistischen Ergebnisse der Drop-Out-Analysen sind in Tabelle 9 bis Tabelle 12 dargestellt.

Tabelle 9: Soziodemographische Patientenmerkmale: Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe

	Gesamt (Verlaufs- Stichprobe) (N = 576)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
Soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>t-Test / χ²-Test /FE-Test</i>
Geschlecht				
Frauen	447 (77.6)	333 (78.5)	114 (75.0)	$\chi^2 (1, n = 576) = 0.806$ $p = .369, \phi = -.037$
Männer	129 (22.4)	91 (21.5)	38 (25.0)	
Alter				
in Jahren (Range: 18 – 77)	43.9 (11.1)	45.3 (10.5)	40.4 (12.0)	$t(574) = -4.651$ $p < .001^{***}, d = 0.47^{**}$
Familienstand/ Partnerschaft				
mit Partner (verheiratet)	254 (45.0)	193 (46.5)	61 (40.9)	$\chi^2 (3, n = 564) = 2.916$ $p = .405, \phi = .072$
mit Partner (ledig)	72 (12.8)	48 (11.6)	24 (16.1)	
ohne Partner (ledig)	82 (14.5)	58 (14.0)	24 (16.1)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	156 (27.7)	116 (28.0)	40 (26.8)	
Schulabschluss				
Hauptschule	133 (23.7)	94 (22.9)	39 (25.8)	$\chi^2 (3, n = 562) = 2.854$ $p = .415, \phi = .071$
Realschule/ Polytechnische Oberschule	231 (41.1)	170 (41.4)	61 (40.4)	
Abitur/Fachabitur	178 (31.7)	135 (32.8)	43 (28.5)	
sonstige oder ohne Schulabschluss	20 (3.6)	12 (2.9)	8 (5.3)	
Berufsabschluss				
noch in Ausbildung	11 (2.1)	8 (2.0)	3 (2.2)	$FE (n = 531) = 4.682$ $p = .455, \phi = .092$
Lehre/Fachschule	299 (56.3)	222 (56.5)	77 (55.8)	
Meister/Techniker	70 (13.2)	52 (13.2)	18 (13.0)	
Fachhochschule/ Universität	96 (18.1)	76 (19.3)	20 (14.5)	
ohne Berufsabschluss	43 (8.1)	27 (6.9)	16 (11.6)	
sonstiges	12 (2.3)	8 (2.0)	4 (2.9)	
Erwerbstätigkeit				
Vollzeit berufstätig	274 (52.6)	209 (54.7)	65 (46.8)	$\chi^2 (2, n = 521) = 3.512$ $p = .173, \phi = .082$
Teilzeit berufstätig	128 (24.6)	93 (24.3)	35 (25.2)	
nicht berufstätig	119 (22.8)	35 (25.2)	39 (28.1)	

	Gesamt (Verlaufs- Stichprobe) (N = 576)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
Soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
monatliches Haushaltsnetto- einkommen				
bis 1 000 Euro	81 (14.1)	59 (13.9)	22 (14.5)	
bis 2 000 Euro	189 (32.8)	130 (30.7)	59 (38.8)	
bis 3 000 Euro	143 (24.8)	108 (25.5)	35 (23.0)	$\chi^2 (4, n = 576) = 5.194$ $p = .268, \phi = .095$
bis 4 000 Euro	72 (12.5)	59 (13.9)	13 (8.6)	
mehr als 4 000 Euro	91 (15.8)	68 (16.0)	23 (15.1)	

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; d : Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

In fast allen soziodemographischen Merkmalen zeigen sich keine signifikanten Unterschiede bzw. Unterschiede mit mindestens kleiner Effektstärke zwischen der Verlaufs-Outcome-Stichprobe und der Drop-Out-Stichprobe zur Katamnese. Lediglich im Alter besteht ein signifikanter Unterschied. Die Patienten, die zur Katamnese nicht mehr teilgenommen haben, waren älter als die teilnehmenden Patienten. Der Unterschied erreicht eine annähernd mittlere Effektstärke (s. Tabelle 9).

Tabelle 10: Sozialmedizinische und klinische Patientenmerkmale: Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe

	Gesamt (Verlaufs- Stichprobe) (N = 576)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
Sozialmedizinische/ klinische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>t-Test / χ^2-Test</i>
Behandlungsdauer				
in Tagen (Range 19 – 60)	42.0 (8.5)	42.2 (8.2)	41.5 (9.2)	$t(574) = -0.814$ $p = .416, d = 0.09$
Arbeitsunfähigkeits-/ Krankheitszeiten (6 Monate vor t0)				
in Wochen (Range 0.2 - 26)	9.8 (9.4)	10.0 (9.5)	9.4 (9.1)	$t(536) = -.576$ $p = .565, d = 0.06$
Psychotherapeutische Vorbehandlung (6 Monate vor t0)				
Ambulante Vorbehandlung	277 (48.5)	205 (48.9)	72 (47.4)	$\chi^2 (1, n = 571) = 0.108$ $p = .742, \phi = .014$
keine ambulante Vorbehandlung	294 (51.5)	214 (51.1)	80 (52.6)	
Stationäre Vorbehandlung	58 (10.1)	42 (10.0)	16 (10.5)	$\chi^2 (1, n = 574) = 0.040$ $p = .841, \phi = -.008$
keine stationäre Vorbehandlung	516 (89.9)	380 (90.0)	54 (89.5)	
Symptombelastung				
HEALTH-49 A t0 (Range 0 – 3.72)	1.41 (0.78)	1.35 (0.76)	1.55 (0.83)	$t(574) = 2.698$ $p = .007^{**}, d = -0.26^+$
HEALTH-49 A t0: unterhalb des Cut-Off- Wertes (0.685)	112 (19.4)	92 (21.7)	20 (13.2)	$\chi^2 (1, n = 576) = 5.210$ $p = .022^*, \phi = -.095$
Lebensqualität				
SF-8 Psych. t0 (Range 7.43 – 61.36)	32.08 (12.16)	32.72 (12.10)	30.24 (12.20)	$t(563) = -2.131$ $p = .034^*, d = -0.20^+$
Selbstwirksamkeits- erwartung				
SWE t0 (Range 10 - 40)	23.48 (6.51)	23.72 (6.51)	22.80 (6.48)	$t(570) = -1.494$ $p = .136, d = -0.14$

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**), und große (***) Effektstärken interpretiert.

Im Hinblick auf die krankheitsbezogenen Merkmale zeigt sich, dass die Patienten der Verlaufs-Outcome-Stichprobe eine geringere Symptombelastung bei Aufnahme in der Klinik aufwiesen als die Patienten, die zum Katamnese-Zeitpunkt nicht mehr teilnahmen. Die mittlere initiale Symptombelastung in der Verlaufs-Outcome-Stichprobe ist mit 1.35 (SD = 0.76) zudem signifikant niedriger als der Mittelwert der entsprechenden Referenzstichprobe

stationär behandelter Psychotherapiepatienten bei Aufnahme [Ein-Stichproben-*t*-Test HEALTH-49 $A_{\text{stat. Psychotherapie}}: M = 1.45 (SD = 0.77); t(423) = -2.597, p = .010^{**}, d = 0.13$] (Rabung et al., 2008). Der Unterschied zur Referenzstichprobe erreicht jedoch nicht die Höhe einer kleinen Effektstärke. Ein entsprechendes Bild zeigt sich bei den psychischen Aspekten der gesundheitsbezogenen Lebensqualität. Patienten der Verlaufs-Outcome-Stichprobe hatten bei Behandlungsbeginn höhere Lebensqualitätswerte als Patienten, die die Studienteilnahme abgebrochen haben. In beiden Skalen erreichen die Unterschiede zwischen Verlaufs-Outcome- und Katamnese-Drop-Out-Stichprobe die Höhe einer kleinen Effektstärke (s. Tabelle 10). Die etwas geringere Ausgangsbelastung in der Verlaufs-Outcome-Stichprobe muss bei der Interpretation der Ergebnisse berücksichtigt werden. In allen anderen sozialmedizinischen und klinischen Merkmalen bestehen keine signifikanten Unterschiede zwischen Verlaufs-Outcome-Stichprobe und der Drop-Out-Stichprobe zur Katamnese.

Tabelle 11: Diagnoseverteilung: Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe

Diagnose	ICD-10 Code	Gesamt (Verlaufs-Stichprobe) (N = 571)	Verlaufs-Outcome-Stichprobe (n = 424)	Katamnese-Drop-Out-Stichprobe (n = 152)	Signifikanztest
		n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Entlassdiagnose (1. - 10.)					
Unipolare Depression	F32, F33, F34.1	385 (67.4)	286 (67.6)	99 (66.9)	$\chi^2 (1, n = 571) = 0.026$ $p = .872, \phi = .007$
Angststörung	F40, F41, F48	105 (18.4)	72 (17.0)	33 (22.3)	$\chi^2 (1, n = 571) = 2.034$ $p = .154, \phi = -.060$
PTBS	F43.1	31 (5.4)	21 (5.0%)	10 (6.8)	$\chi^2 (1, n = 571) = 0.686$ $p = .408, \phi = -.035$
Anpassungsstörung	F43.2	103 (18.0)	86 (20.3)	17 (11.5)	$\chi^2 (1, n = 571) = 5.801$ $p = .016^*, \phi = .101^+$
Somatoforme Störung	F45	81 (14.2)	63 (14.9)	18 (12.2)	$\chi^2 (1, n = 571) = 0.672$ $p = .412, \phi = .034$
Essstörung	F50	96 (16.8)	59 (13.9)	37 (25.0)	$\chi^2 (1, n = 571) = 9.575$ $p = .002^{**}, \phi = -.129^+$
Persönlichkeitsstörung	F60 – F61	36 (6.3)	24 (5.7)	12 (8.1)	$\chi^2 (1, n = 571) = 1.100$ $p = .294, \phi = -.044$
Andere psych. Störung		79 (13.8)	51 (12.1)	28 (18.9)	$\chi^2 (1, n = 571) = 4.331$ $p = .037^*, \phi = -.087$
Adipositas	E66	151 (26.4)	107 (25.3)	44 (29.7)	$\chi^2 (1, n = 571) = 1.108$ $p = .292, \phi = -.044$
Erste Entlassdiagnose					
Unipolare Depression	F32, F33, F34.1	325 (57.2)	247 (58.7)	78 (53.1)	FE (n = 568) = 22.079

Diagnose	ICD-10 Code	Gesamt (Verlaufs- Stichprobe) (N = 571)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
		n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Angststörung	F40, F41, F48	42 (7.4)	26 (6.2)	16 (10.9)	$p = .004^{**}$, $\phi = .197^+$
PTBS	F43.1	11 (1.9)	9 (2.1)	2 (1.4)	
Anpassungs- störung	F43.2	90 (15.8)	76 (18.1)	14 (9.5)	
Somatoforme Störung	F45	21 (3.7)	17 (4.0)	4 (2.7)	
Essstörung	F50	41 (7.2)	23 (5.5)	18 (12.2)	
Persönlichkeitsstö- rung	F60 – F61	7 (1.2)	5 (1.2)	2 (1.4)	
Andere psych. Störung		9 (1.6)	4 (1.0)	5 (3.4)	
Adipositas	E66	22 (3.9)	14 (3.3)	8 (5.4)	
Psychische Komorbidität					
Anzahl der F-Diagnosen (Range 0 – 5)		1.6 (0.9)	1.6 (0.8)	1.7 (0.9)	$t(569) = 1.814$ $p = .070$, $d = -0.13$
Mehr als eine F-Diagnose		317 (55.5)	220 (52.0)	97 (65.5)	$\chi^2(1, n = 571) = 8.129$ $p = .004^{**}$, $\phi = -.119^+$

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**), und große (***) Effektstärken interpretiert.

In der Verteilung der Entlassdiagnosen, bei der alle vergebenen Diagnosen berücksichtigt werden, zeigen sich Unterschiede in den *Anpassungsstörungen* und den *Essstörungen*. In der Verlaufs-Outcome-Stichprobe befinden sich mehr Patienten mit Anpassungsstörung als in der Drop-Out-Stichprobe (20% vs. 12%), während die Patienten mit einer Essstörung stärker bei den Studienabbrechern repräsentiert sind. Auch Patienten mit einer anderen als den sieben hier aufgeführten Diagnosegruppen aus dem F-Kapitel des ICD-10, zusammengefasst unter *Andere psychische Störung*, sind stärker in der Drop-Out- als in der Verlaufs-Outcome-Stichprobe vertreten. Hinsichtlich der ersten Entlassdiagnose ist zwischen den Stichproben ebenfalls ein signifikanter Unterschied mit kleiner Effektstärke zu finden. Die Verlaufs-Outcome-Stichprobe ist im Vergleich zur Drop-Out-Stichprobe durch einem höheren Anteil an Patienten mit unipolaren *Depressionen* (59% vs. 53%) und *Anpassungsstörungen* (18% vs. 10%) und einem geringeren Anteil an Patienten mit *Angststörungen* (6% vs. 11%) und *Essstörungen* (6% vs. 12%) gekennzeichnet. In der Verlaufs-Outcome-Stichprobe befinden sich weniger Patienten, die mehr als eine F-Diagnose haben, als in der Drop-Out-

Stichprobe (52% vs. 66%). Die mittlere Anzahl an F-Diagnosen unterscheidet sich aber nicht zwischen den beiden Stichproben.

Die Verteilung der ersten Entlassdiagnose über die verschiedenen Studienstichproben ist ausgehend von der Einschlussstichprobe bis zur Verlaufs-Outcome-Stichprobe mit den jeweiligen Nicht-Teilnehmer- und Drop-Out-Stichproben in Abbildung 6 dargestellt. Hier wird noch einmal deutlich, dass die Verteilung der vier Hauptdiagnosegruppen (Erstdiagnose) *unipolare Depressionen*, *Anpassungsstörungen*, *Angststörungen* und *Essstörungen* zwar über die Stichproben hinweg leicht variieren, die Verteilung in den beiden Analyse-Stichproben (Verlaufsstichprobe und Verlaufs-Outcome-Stichprobe) sich jedoch nicht maßgeblich voneinander und von der Verteilung der Einschlussstichprobe unterscheiden.

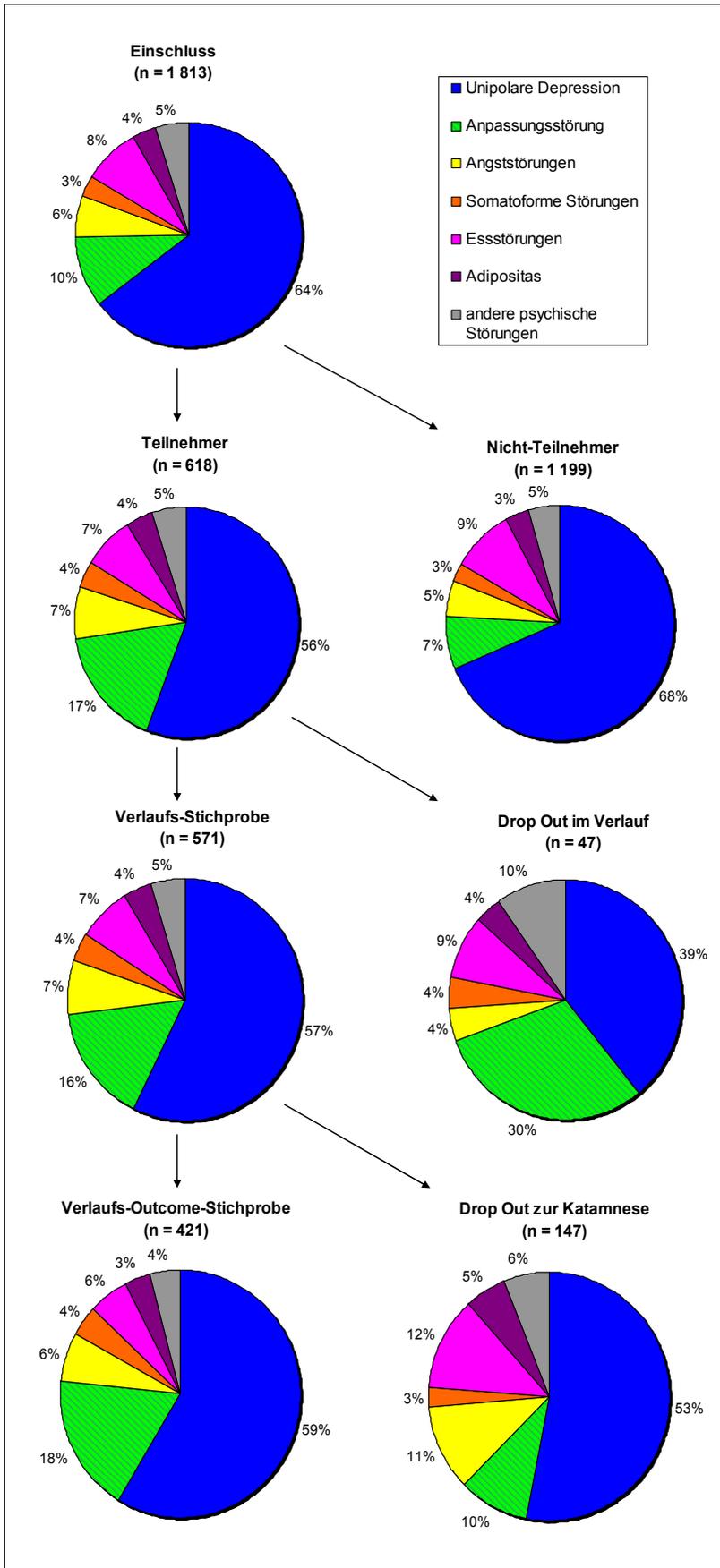


Abbildung 6: Verteilung der ersten Entlassdiagnosen über die Analyse- und Drop-Out-Stichproben

Die Drop-Out-Raten zur Katamnese unterscheiden sich signifikant zwischen den Kliniken. Klinik 1 (36%; 67 von 186) und Klinik 3 (28%, 43 von 153) weist einen deutlich höheren Drop-Out auf als Klinik 2 (18%, 42 von 237). Die Anteile der Patienten aus den drei Kliniken sind in der Verlaufs-Outcome-Stichprobe wie in der Katamnese-Drop-Out-Stichprobe ungleich verteilt (s. Tabelle 12).

Tabelle 12: Vergleich der Drop-Out-Raten zur Katamnese zwischen den drei Kliniken

Klinik	Gesamt (Verlaufs- Stichprobe) (N = 576)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
	n (%)	n (%)	n (%)	χ^2 -Test
Klinik 1	186 (100.0)	119 (64.0)	67 (36.0)	$\chi^2 (2, n = 576) = 18.282$ $p < .001^{***}$, $\phi = .178^+$
Klinik 2	237 (100.0)	195 (82.3)	42 (17.7)	
Klinik 3	153 (100.0)	110 (71.9)	43 (28.1)	

χ^2 : Chi²-Test; p : Signifikanz; ϕ : Effektstärke Phi. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Insgesamt zeigt sich, dass die Patienten, die an der Katamnese-Erhebung teilnahmen, bedeutend älter waren als die Patienten, die zur Katamnese nicht mehr teilnahmen. Zudem wiesen die Patienten der Verlaufs-Outcome-Stichprobe zu Beginn der Behandlung eine geringere Symptombelastung und eine höhere Lebensqualität auf als die Patienten der Drop-Out-Stichprobe. In der Verlaufs-Outcome-Stichprobe waren mehr Patienten mit der Erstdiagnose einer unipolaren Depression und Anpassungsstörung zu finden und weniger Patienten mit Angststörungen und Essstörungen als in der Drop-Out-Stichprobe. Die Drop-Out-Raten unterschieden sich signifikant zwischen den drei Kliniken. Die potentiellen Selektionseffekte, vor allem hinsichtlich der initialen Belastung, müssen bei der Interpretation der Ergebnisse berücksichtigt werden.

5.2 Ergebnisse der explorativen Verlaufsanalyse (Fragestellung 1 A)

5.2.1 Identifikation der Verlaufsklassen

Zur Identifikation latenter Verlaufsklassen kamen Verlaufsanalysen mittels Growth Mixture Modeling (GMM) zum Einsatz (s. Abschnitt 4.5.2). In Tabelle 13 sind die identifizierten Klassen mit steigender Klassenzahl aufgeführt. Neben der Klassenanzahl und -größe werden die Log Likelihood, die Anzahl freier Parameter, das Informationskriterium BIC, die Tests auf signifikant besserer Modellpassung (ab 2-Klassenlösung) sowie die Entropie der jeweiligen Klassenlösung angegeben. Diese Parameter ermöglichen den Vergleich der Modellgüte für die Modelle mit verschiedener Anzahl an Klassen.

Tabelle 13: Klassengrößen und Modellgütekriterien der Klassenlösungen 1 - 6 der Initialphase und 6-Klassenlösung der Validierungsphase

	Log Likelihood	FP	BIC	LMR-LRT	BLRT	Entropie	Klassengröße in %
1 Klasse	-1 798.930	18	3 712.270	--	--	--	100
2 Klassen	-1 748.248	22	3 636.330	0.0015	< 0.0001	0.828	85; 15
3 Klassen	-1 732.379	26	3 630.016	0.3556	< 0.0001	0.857	85; 13; 2
4 Klassen	-1 720.477	30	3 631.637	0.5996	0.0128	0.848	82; 12; 4; 1
5 Klassen	-1 707.511	34	3 631.129	0.5705	< 0.0001	0.778	73; 13; 9; 4; 2
6 Klassen	-1 694.251	38	3 630.034	< 0.0001	< 0.0001	0.818	71; 13; 11; 3; 2; 1
6 Klassen Validierung	-1 698.173	38	3 637.878	--	--	0.801	73; 12; 10; 4; 2; 1

FP: Freie Parameter; BIC: Bayesian Information Criterion; LMR-LRT: Lo-Mendell-Rubin Adjusted Likelihood Ratio Test; BLRT: Bootstrapped Likelihood Ratio Test

In Abbildung 7 und Abbildung 8 sind die BIC-Werte und die Log Likelihood-Werte über die 1- bis 6-Klassenlösungen graphisch dargestellt.

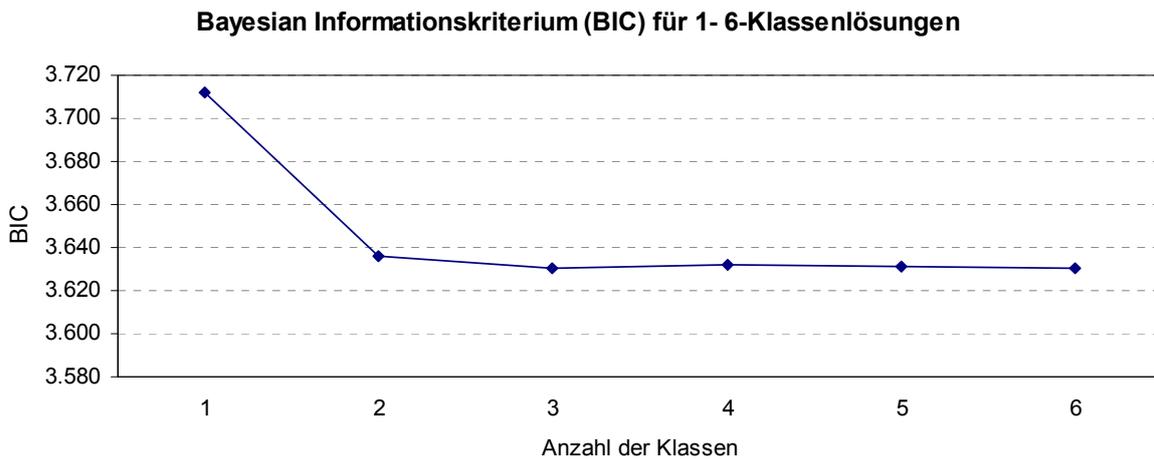


Abbildung 7: Graphische Darstellung des BIC für 1- bis 6-Klassenlösungen

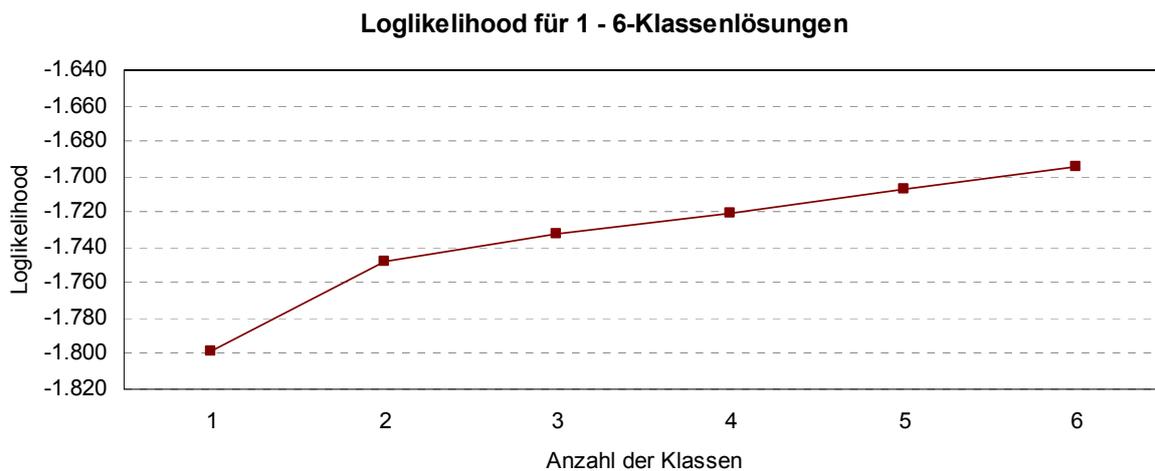


Abbildung 8: Graphische Darstellung der Log Likelihood-Werte für 1- bis 6-Klassenlösungen

Zur statistischen Prüfung der besten Klassenlösungen, d.h. desjenigen Modells, in dem die Modellierung der Daten den tatsächlichen Daten am besten entspricht, werden mehrere Parameter herangezogen. Es wurden Modelle in aufsteigender Klassenanzahl mit bis zu 6 Klassen berechnet. Eine weitere Differenzierung, d.h. die Prüfung einer 7-Klassenlösung wurde aus inhaltlichen Gründen nicht vorgenommen. Aus den vorherigen Klassenlösungen wird ersichtlich, dass sich neben einer großen Klasse, die relativ konstant über die verschiedenen Modelle hinweg identifiziert werden kann, eine Aufspaltung mehrerer kleiner Klassen stattfindet. Eine weitere Aufdifferenzierung würde die einzelnen Klassen zu klein und somit nicht mehr zielführend hinsichtlich der Aussagen über prototypische Verläufe werden lassen.

Das BIC ist bei der 6-Klassenlösung am niedrigsten, betrachtet man das BIC allerdings hinsichtlich des Ellenbogen-Kriteriums, welches eine sprunghafte Abnahme der Werte anzeigt (s. Abbildung 7), zeigt sich, dass die 2-Klassenlösung vorgezogen werden sollte. Ähnliches gilt für die Log Likelihood, bei der der niedrigste Wert bei der 6-Klassenlösung zu finden ist, ein größerer Sprung in der Abnahme der Werte aber bei der 2-Klassenlösung auftaucht. Dahingegen weist das Verfahren LMRT, bei dem die nächst höhere Klassenlösung mit der vorherigen Klassenlösung hinsichtlich der Passung der Daten auf statistische Signifikanz getestet wird, auf eine Überlegenheit der 6-Klassenlösung hin. Die Entropie, die die Zuordnungsgüte der Patienten zu den Klassen quantifiziert, liegt bei allen Klassenlösungen in einem hohen Bereich ($> .75$). Aus den statistischen Kennwerten zeigt sich demnach eine deutliche Tendenz einer optimalen Passung des Modells bei der 2-Klassenlösung oder bei der 6-Klassenlösung. Aus inhaltlichen Überlegungen erscheint es sinnvoll, die 6-Klassenlösung heranzuziehen, weil hier ein deutlich höherer Informationsgewinn erzielt wurde. Um die statistische Stabilität der 6-Klassenlösung

abzusichern, wurde zusätzlich in einer Validierungsphase geprüft, ob sich die identifizierten sechs Klassen bei vorgegebenen Parametern der jeweils zuvor identifizierten Klassen replizieren lassen. Die Ergebnisse der Validierungsphase sind ebenfalls in Tabelle 13 dargestellt. Bei der aus der Validierungsphase erzielten 6-Klassenlösung konnte der niedrigste Log Likelihood-Wert repliziert werden.

In Abbildung 9 ist die Verteilung der Patienten über alle modellierten Klassenlösungen in Form eines sogenannten *Class Evolution Tree* graphisch dargestellt (Kriston, Melchior, Hergert, Bergelt, Watzke, Schulz et al., 2011). In den Feldern sind die absoluten und relativen Häufigkeiten der zur jeweiligen Klasse zugeordneten Patienten angegeben. Der Prozentwert unterhalb des Feldes zeigt den Anteil der Patienten an, die in der darunterliegenden Klasse der nächsten Klassenlösung wiederzufinden sind (Zielproportion). Die Prozentzahl oberhalb des Feldes gibt den Anteil an Patienten an, der der darüber liegenden Klasse ($k-1$ Klassenlösen) entstammt (Quellproportion). Auf der untersten Ebene ist die präferierte 6-Klassenlösung dargestellt.

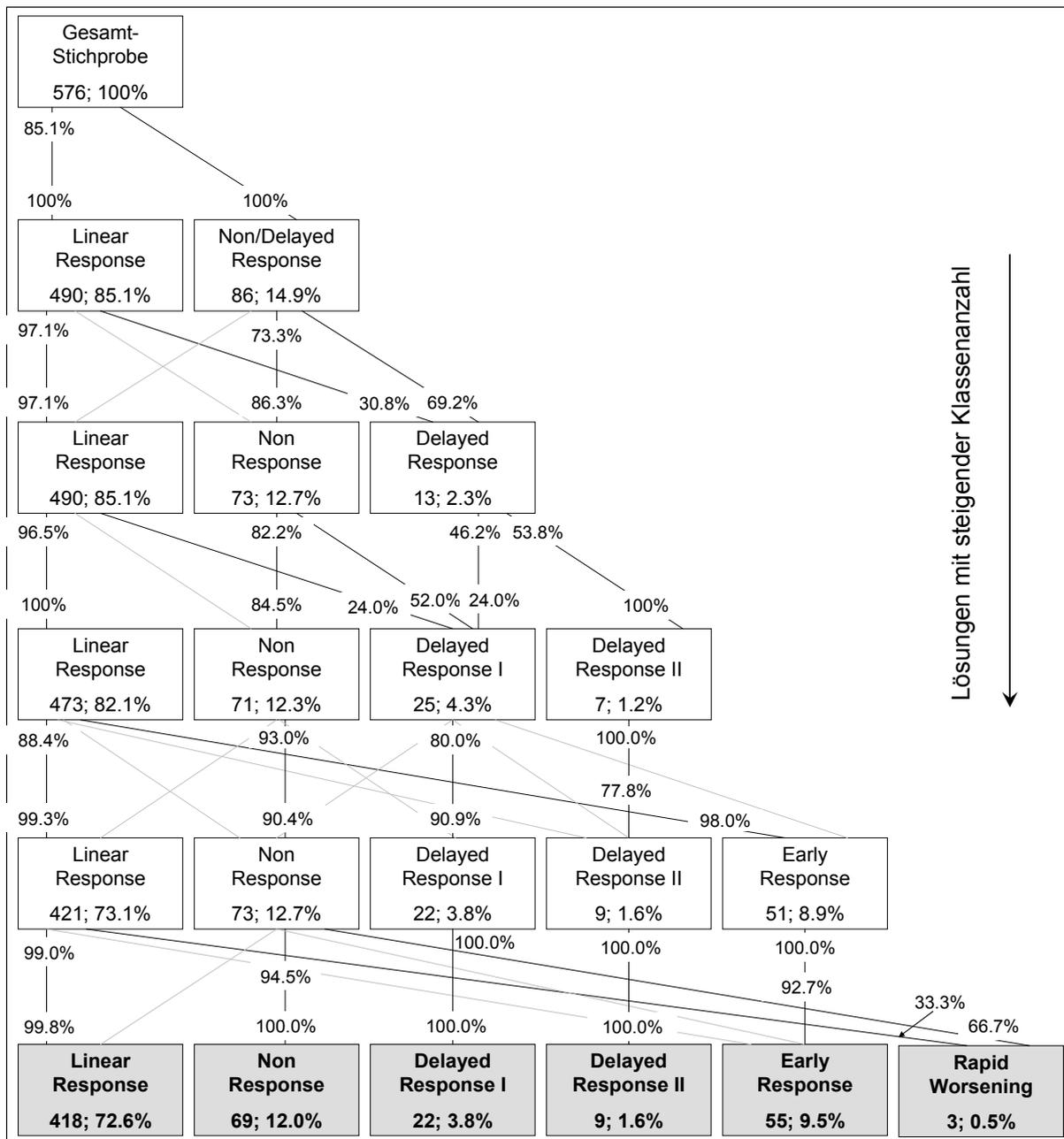


Abbildung 9: Class Evolution Tree – Verteilung und Patientenfluss über die sechs modellierten Klassenlösungen (modifiziert nach Kriston et al., 2011)

5.2.2 Beschreibung der Verlaufsklassen

Nach der statistischen Absicherung der 6-Klassenlösung und der inhaltlichen Einschätzung des Informationsgewinns und der Interpretierbarkeit der einzelnen Klassen wurde die Entscheidung für die 6-Klassenlösung getroffen. Die Kurvenparameter Intercept, linearer und quadratischer Slope sind mit den jeweiligen Standardfehlern für die sechs Klassen in Tabelle 14 dargestellt. Abbildung 10 zeigt die graphische Darstellung der mittleren Verlaufskurven aller sechs Klassen.

Tabelle 14: Kurvenparameter der sechs Verlaufsklassen

Klasse	n	Intercept (SE)	linearer Slope (SE)	quadratischer Slope (SE)
1	9	2.632 (0.272)	0.048 (0.130)	-0.061 (0.019)
2	3	0.875 (0.642)	-0.464 (0.351)	0.232 (0.086)
3	69	1.863 (0.158)	-0.053 (0.045)	0.007 (0.006)
4	418	1.008 (0.061)	-0.119 (0.023)	0.006 (0.003)
5	55	2.067 (0.131)	-0.506 (0.087)	0.044 (0.011)
6	22	2.651 (0.356)	0.083 (0.145)	-0.034 (0.016)

SE: Standardfehler (standard error)

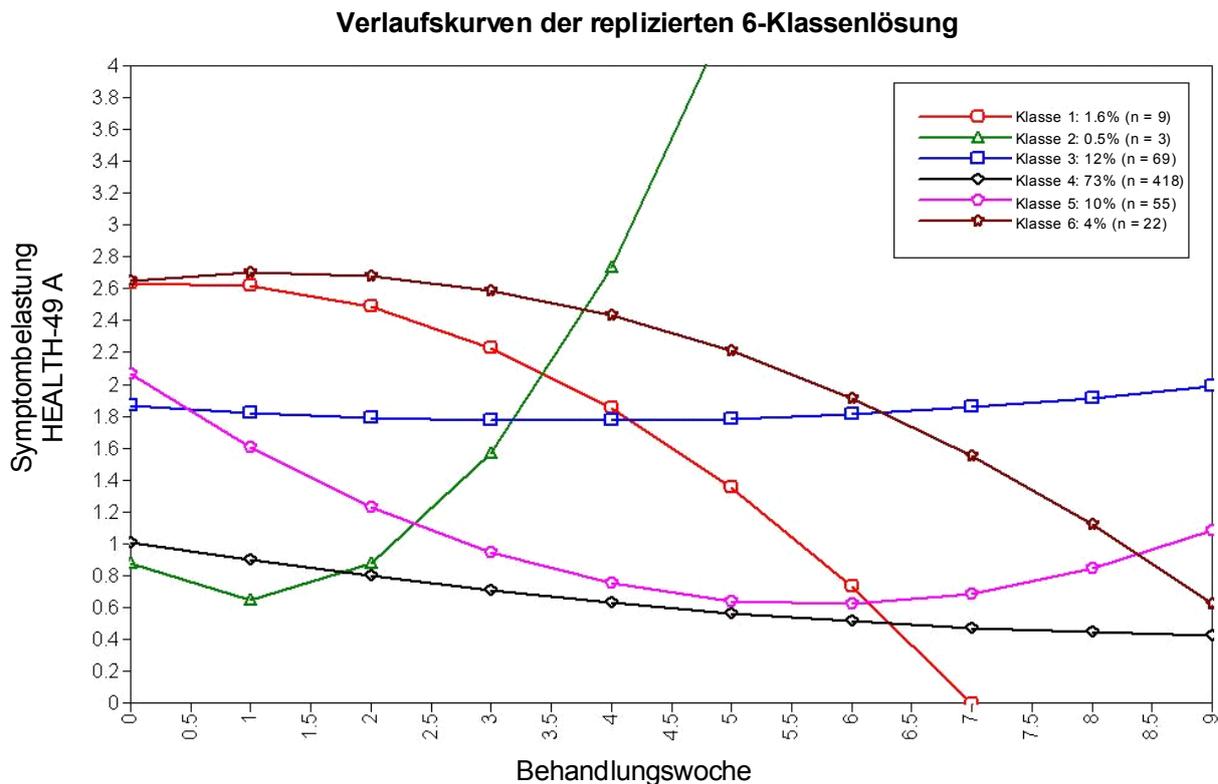


Abbildung 10: Verlaufskurven der sechs Verlaufsklassen

Bei Betrachtung der Verlaufskurven der einzelnen Klassen muss berücksichtigt werden, dass diese keine Endpunkte aufweisen, sodass das Ende der Behandlung bzw. die Behandlungsdauer aus dieser Abbildung nicht hervorgeht.

Patienten der Klasse 1 (braun) und Klasse 6 (rot) zeichnen sich durch eine ähnlichen Verlauf mit hoher Ausgangsbelastung und einer verzögerten Verbesserung im Verlauf aus. In den ersten Behandlungswochen treten in beiden Klassen erst einmal nur geringe Veränderungen auf. Nach einem gewissen Zeitraum von etwa 3 Wochen weisen diese jedoch einen deutlichen Abfall der Symptombelastung in Richtung Verbesserung auf. Patienten der Klasse 3 (blau) lassen sich durch eine mittlere bis hohe Ausgangsbelastung und eine anscheinend geringe bis gar keine Veränderung im Verlauf der Behandlung beschreiben. Die Klasse 5

(magenta) ist durch eine ähnlich mittlere bis hohe Ausgangsbelastung gekennzeichnet, hier kommt es jedoch schon im frühen Verlauf der Behandlung zu einer deutliche Senkung der Symptombelastung und demnach zu einer frühen Verbesserung. Nach der Phase der frühen Verbesserung bleibt die Symptombelastung auf einem konstant niedrigen Niveau. Der augenscheinliche Wiederanstieg der Verlaufskurve nach der sechsten bis siebten Woche ist bei genauer Betrachtung aller zugehörigen Einzelverläufe lediglich durch den fehlenden Endpunkt der Kurve begründet (s. Abbildung 15; S. 116). Nur wenige Patienten dieser Klasse werden länger als 6 Wochen behandelt, eine 7-wöchige Behandlungsdauer wird in dieser Klasse nicht überschritten. Die vierte Klasse (schwarz), die auch die größte Klasse darstellt, ist durch eine mittlere bis niedrige Ausgangsbelastung und eine stetige, aber eher geringe Verbesserung über den Verlauf gekennzeichnet. Die Patienten der Klasse 2 (grün), der kleinsten Klasse, weisen einen ungewöhnlichen Verlauf auf: Bei einer mittleren bis niedrigen Symptombelastung zu Behandlungsbeginn kommt es bei diesen Patienten nach einem kurzen Zeitraum von etwa 2 Wochen zu einem deutlichen Anstieg der Belastung. Allerdings muss auch hier beachtet werden, dass die Verlaufsmessung nach 3 bzw. 4 Wochen endet (s. Abbildung 12), was entweder auf eine kurze Behandlungsdauer oder das Fehlen weiterer Verlaufsmesszeitpunkte zurückzuführen ist. In den folgenden Abbildungen (Abbildung 11 bis Abbildung 16) sind die sechs Klassen im Einzelnen mit den beobachteten Verläufen dargestellt. Die jeweils farbige Kurve zeigt die geschätzte mittlere Verlaufskurve der Klasse, die schwarzen unterbrochenen Verläufe sind die beobachteten Einzelverläufe aller der Klasse zugeordneten Patienten. Anhand der Einzelverläufe kann auch die jeweilige Dauer der Behandlung, bzw. genauer die Anzahl der vorhandenen Verlaufsmesszeitpunkte durch das Ende der Einzelverlaufskurve abgelesen werden, sodass einer Fehlinterpretation der mittleren geschätzten Kurve, die keinen Endpunkt besitzt, vorgebeugt wird.

Verläufe der Klasse 1 (n = 9)

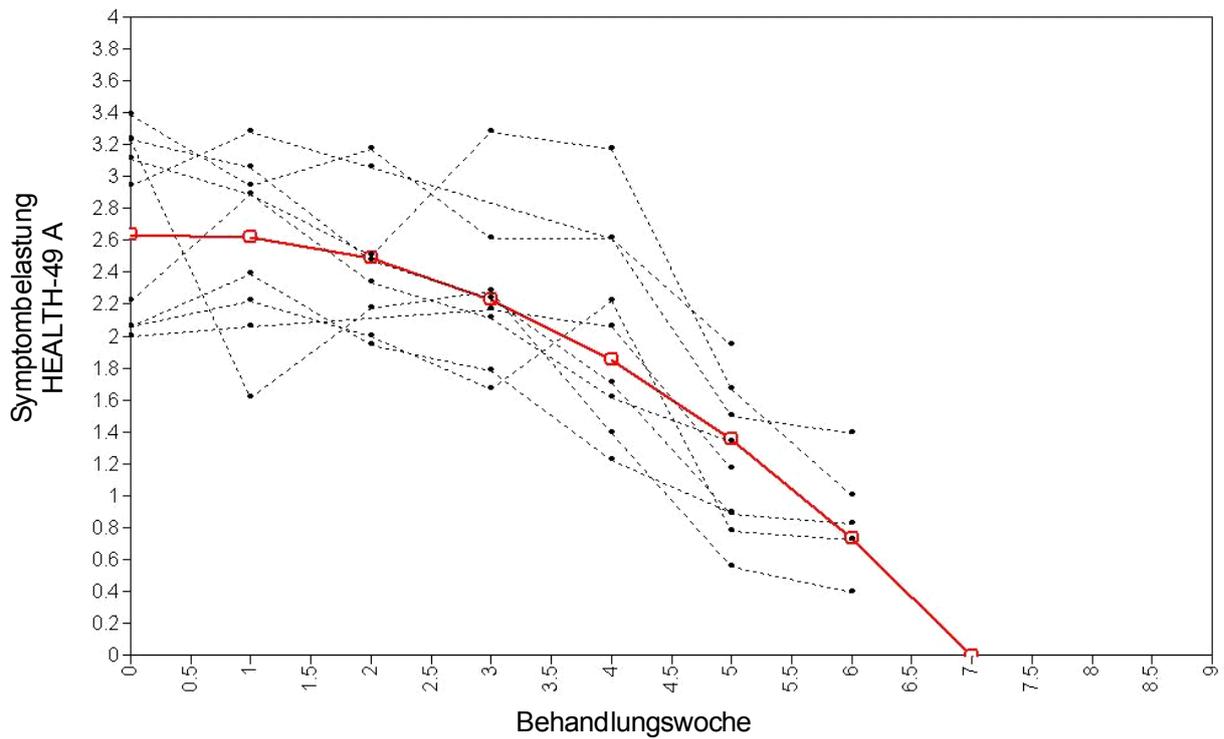


Abbildung 11: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 1 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten (n = 9)

Verläufe der Klasse 2 (n = 3)

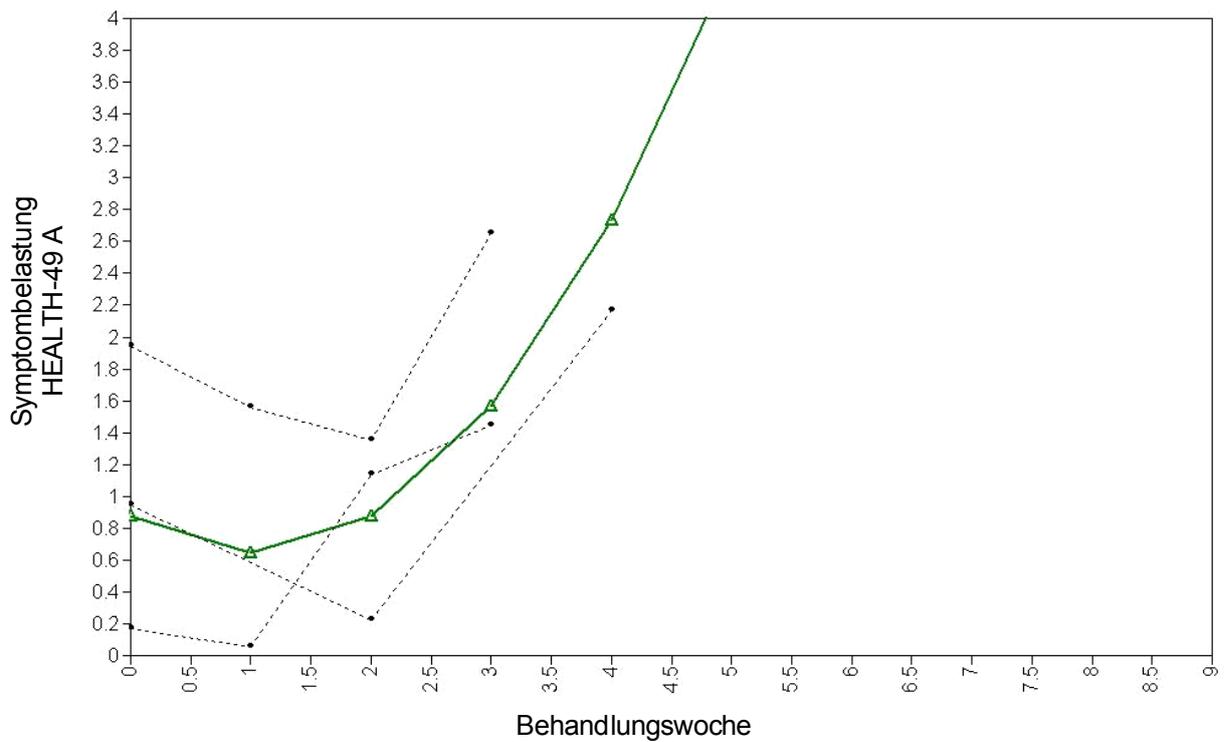


Abbildung 12: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 2 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten (n = 3)

Verläufe der Klasse 3 (n = 69)

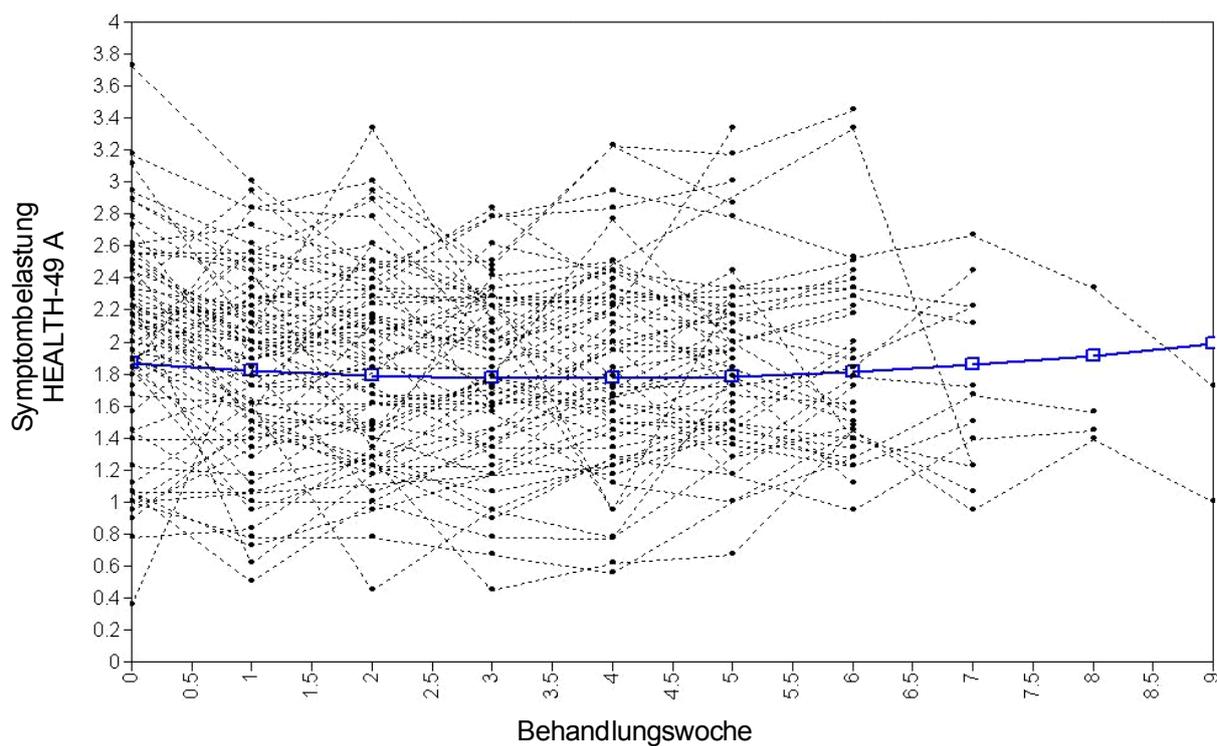


Abbildung 13: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 3 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 69$)

Verläufe der Klasse 4 (n = 418)

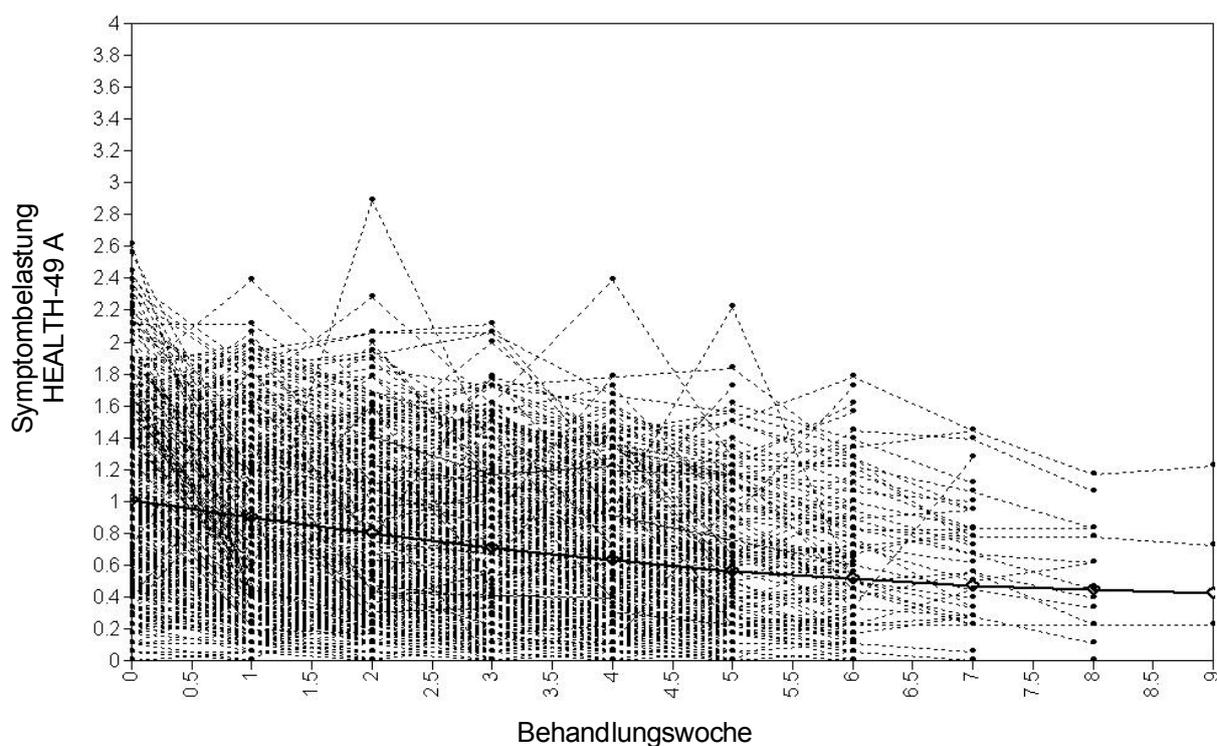


Abbildung 14: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 4 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 418$)

Verläufe der Klasse 5 (n = 55)

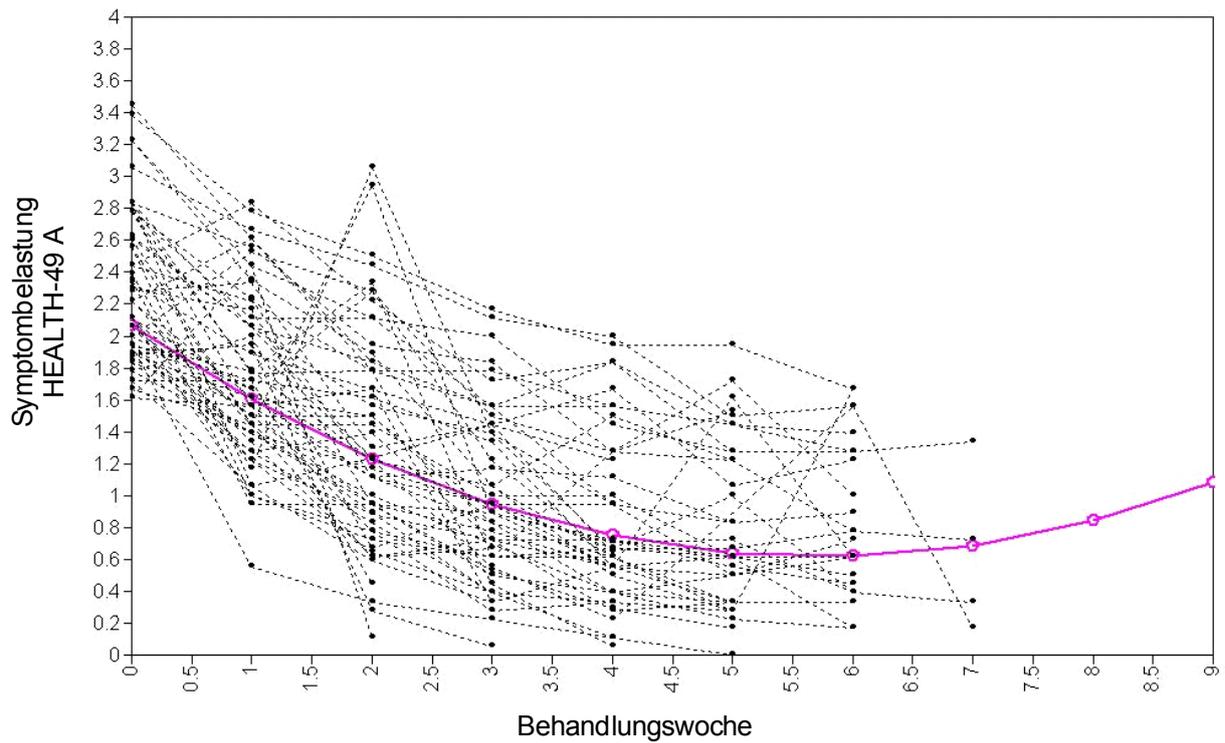


Abbildung 15: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 5 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten (n = 55)

Verläufe der Klasse 6 (n = 22)

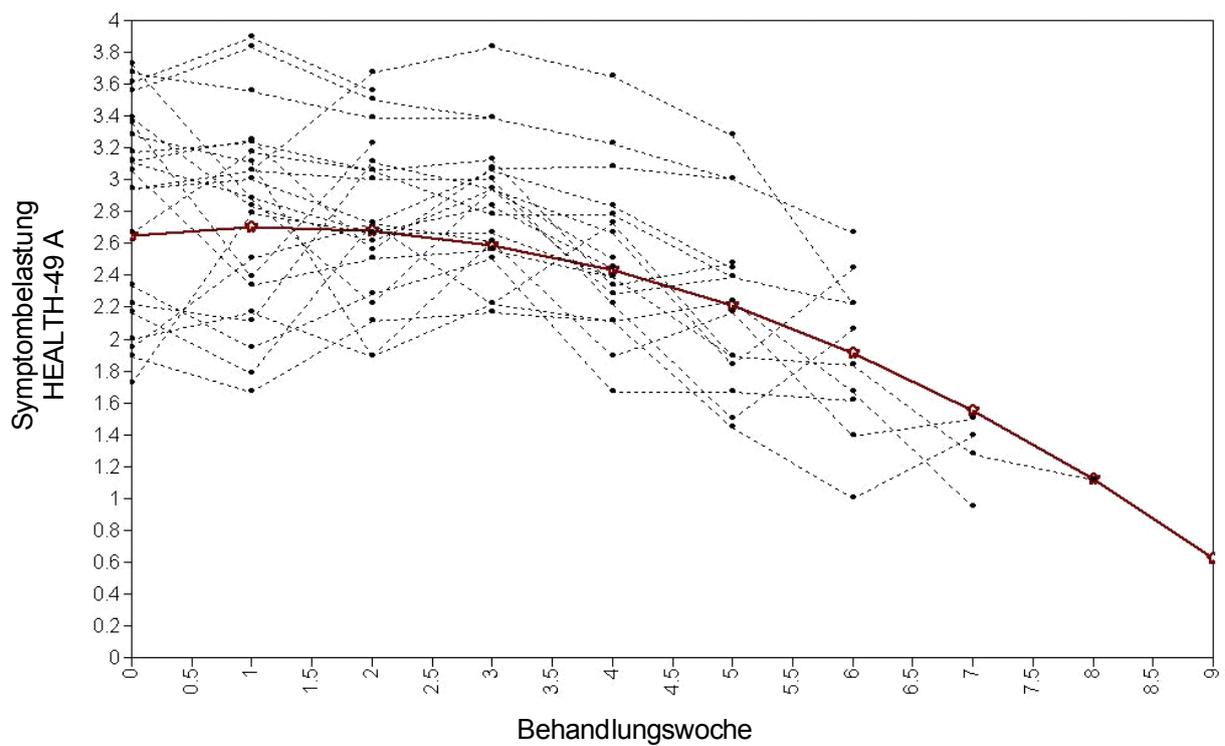


Abbildung 16: 6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 6 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten (n = 22)

5.2.3 Inhaltliche Modifikation der Verlaufsklassen

Bei der Betrachtung der Einzelverläufe fällt zum einen auf, dass neben vielen Verläufen, die augenscheinlich relativ gut zu den Klassen zugeordnet wurden, einige „gezackte“ Verläufe auftreten, deren Verlaufsformen stärker von der geschätzten mittleren Klassenkurve abweichen. Zum anderen wird die große Varianz der Anzahl der zu einer Klasse zugeordneten Patienten deutlich. Die Klasse 2 enthält nur 3 Fälle, die zwar einen auffälligen Verlauf aufweisen, aber durch die geringe Anzahl eher als Extremfälle angesehen werden können. Diese Klasse soll in den weiteren Analysen nicht als inhaltliche Klasse berücksichtigt werden. Die mehr oder weniger gute Zuordnung zu den Klassen kann über die Klassenzuordnungswahrscheinlichkeit quantifiziert werden. Um die Genauigkeit der Zuordnung und somit die Homogenität innerhalb einer Klasse zu erhöhen, werden alle Patienten, deren höchste Klassenwahrscheinlichkeit weniger als 50% beträgt, aus der jeweiligen Klasse ausgeschlossen (s. Tabelle 15). Über alle Klassen hinweg werden insgesamt 19 Patienten (3%) aufgrund einer zu ungenauen Zuordnung aus den einzelnen Klassen entfernt. Diese 19 Patienten, sowie die 3 Patienten der Klasse 2, werden in einer Restklasse zusammengefasst, die im Folgenden nicht inhaltlich interpretiert wird.

Tabelle 15: Patienten mit einer maximalen Klassenzuordnungswahrscheinlichkeit von weniger als 50%

Klasse	<i>n</i> Ausschluss	%	<i>n</i> Klasse alt	<i>n</i> Klasse neu
1	1	11.1%	9	8
2	0	0%	3	3
3	8	11.6%	69	61
4	7	1.3%	418	411
5	2	3.6%	55	53
6	1	4.5%	22	21
Gesamt	19	3.3%	576	557

Zwei der sechs identifizierten Klassen weisen einen relativ ähnlichen Verlauf auf. Patienten der Klasse 1 und Klasse 6 sind durch einen hohen Ausgangswert der Symptombelastung und eine verzögerte Verbesserung gekennzeichnet. Etwa bis zur 2. bis 3. Woche findet keine deutliche Veränderung der Symptomatik statt, was sich jedoch nach diesem Zeitraum ersichtlich ändert, die Symptombelastung fällt bis zum Ende der Behandlung deutlicher ab. Obwohl die verzögerte Verbesserung bei Klasse 1 etwas früher einsetzt als bei Klasse 6, scheint es, auch aufgrund der geringen Klassengrößen ($n = 9$ und $n = 22$), inhaltlich sinnvoller, die beiden Klassen nicht als distinkte Klassen zu betrachten, sondern zu einer Klasse zusammenzufassen. Aus den genannten inhaltlichen Modifikationen resultieren die in Tabelle 16 aufgeführten Klassen.

Tabelle 16: Inhaltlich modifizierte Klassen: Verlaufsgruppen

Klasse	Benennung	Charakteristika	Größe: n (%)
1	Linear Response	niedrige Ausgangsbelastung, stetige, lineare Verbesserung	411 (71%)
2	Non-Response	mittlere Ausgangsbelastung, geringe bis keine Verbesserung	61 (11%)
3	Early Response	mittlere Ausgangsbelastung, frühe Verbesserung	53 (9%)
4	Delayed Response	hohe Ausgangsbelastung, verzögerte Verbesserung	29 (5%)
5	Atypical Course	Atypische Verläufe	22 (4%)

Die Benennungen der Klassen sind angelehnt an die in der Literatur vorhandenen Bezeichnungen für unterschiedliche Verläufe in der psychotherapeutischen Behandlung (vgl. Abschnitt 2.5.4). Dabei werden die Begriffe *Early Response*, *Delayed Response* und *Non-Response* in direktem Bezug zur vorhandenen Terminologie gewählt. Die Begriffe *Linear Response* und *Atypical Course* sind Beschreibungen, die aus keiner theoretisch-konzeptuellen Basis hervorgehen, der Einheitlichkeit halber werden hier auch englische Begriffe verwendet. Zur klaren Abgrenzung der statistisch identifizierten von den inhaltlich modifizierten Klassen werden letztere in den folgenden Analysen sowie in der Diskussion der Ergebnisse nicht mehr als Klassen, sondern als Verlaufsgruppen bezeichnet. Aufgrund der explorativen Vorgehensweise bei der Identifikation der Verlaufsgruppen werden diese zur Abgrenzung der im folgenden Abschnitt (5.3) beschriebenen Gruppen des Weiteren *explorative Verlaufsgruppen* genannt.

Tabelle 17: Absolute und relative Häufigkeit der explorativen Verlaufsgruppen

Explorative Verlaufsgruppe	n	%	95% CI ²⁴	% kumuliert
Linear Response	411	71.4	[67.7; 75.1]	71.4
Early Response	53	9.2	[6.8; 11.6]	80.6
Delayed Response	29	5.0	[3.3; 6.8]	85.6
Non-Response	61	10.6	[8.1; 13.1]	96.2
Atypical Course	22	3.8	[2.3; 5.4]	100
Gesamt	576	100		

CI: Konfidenzintervall

²⁴ Das Konfidenzintervall der relativen Häufigkeiten wird mit der Approximation durch die Normalverteilung geschätzt mit der Annahme einer binomialverteilten Grundgesamtheit. Die Formel zur Berechnung des Konfidenzintervalls bei einer 95%igen Sicherheit lautet:

$$[p - 1.96 \cdot \sqrt{(p \cdot (1-p)/n)}; p + 1.96 \cdot \sqrt{(p \cdot (1-p)/n)}]$$

wobei p der Punktschätzer des relativen Anteils an der Stichprobe und n die Stichprobengröße ist (Altman, Trevor, Gardner & Machin, 2000).

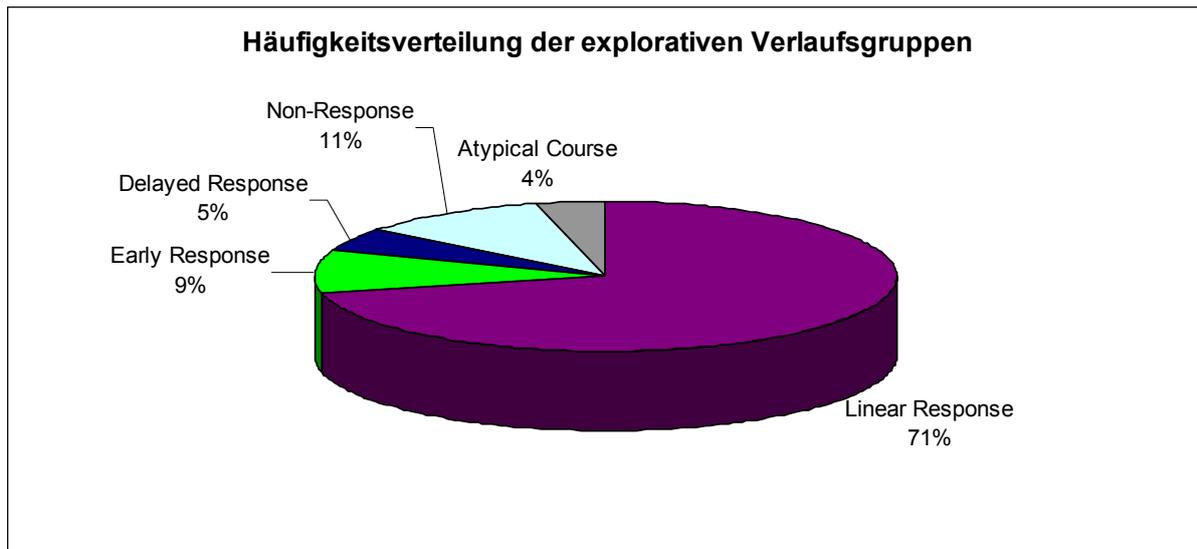


Abbildung 17: Häufigkeitsverteilung der explorativen Verlaufsguppen

Wie aus Tabelle 17 und Abbildung 17 ersichtlich, resultiert die folgende Verteilung der Verlaufsguppen: Die größte Gruppe mit 71% der Stichprobe bilden die 411 Patienten mit einer Linear Response. Die zweitgrößte Gruppe mit 11% der Stichprobe ($n = 61$) stellt die Verlaufsguppe mit der Non-Response dar, gefolgt von der drittgrößten Gruppe der Patienten mit einer Early Response ($n = 53$; 9%). Die kleinste inhaltlich beschreibbare Gruppe bilden die 29 Patienten (5%) mit einer Delayed Response. Die verbleibenden 22 Patienten (4%), die nicht zuverlässig genug zu einer der vier inhaltlichen Verlaufsguppen zugeordnet werden konnten, wurden in der Restkategorie Atypical Course zusammengefasst. Diese Gruppe wird bei den weiteren Analysen berücksichtigt, nicht aber inhaltlich interpretiert.

Die deskriptive Darstellung der Merkmale der einzelnen explorativen Verlaufsguppen ist Teil der zweiten Fragestellung und wird in Abschnitt 5.6.1 dargestellt.

5.3 Ergebnisse der theoriegeleiteten Verlaufsanalyse (Fragestellung 1 B)

Im Folgenden werden die Ergebnisse zu Fragestellung 1 B und *Hypothese 1*, in der die Verteilung von theoretisch definierten Verlaufsguppen fokussiert wird, beschrieben. Die Verteilung der Patienten auf die theoriegeleiteten Gruppen, erstellt nach den in Abschnitt 4.3.4 beschriebenen Kriterien, ist in Tabelle 18 und Abbildung 18 dargestellt.

Tabelle 18: Absolute und relative Häufigkeit der theoriegeleiteten Verlaufsguppen

Theoriegeleitete Verlaufsguppe	<i>n</i>	%	95% CI	% kumuliert
Early Response	287	49.8	[45.7; 53.8]	49.8
Late Response	56	9.7	[7.3; 12.1]	59.5
Non-Response	139	24.1	[20.6; 27.6]	83.6
n/a	94	16.3	[13.3; 19.3]	100
Gesamt	576	100		

CI: Konfidenzintervall

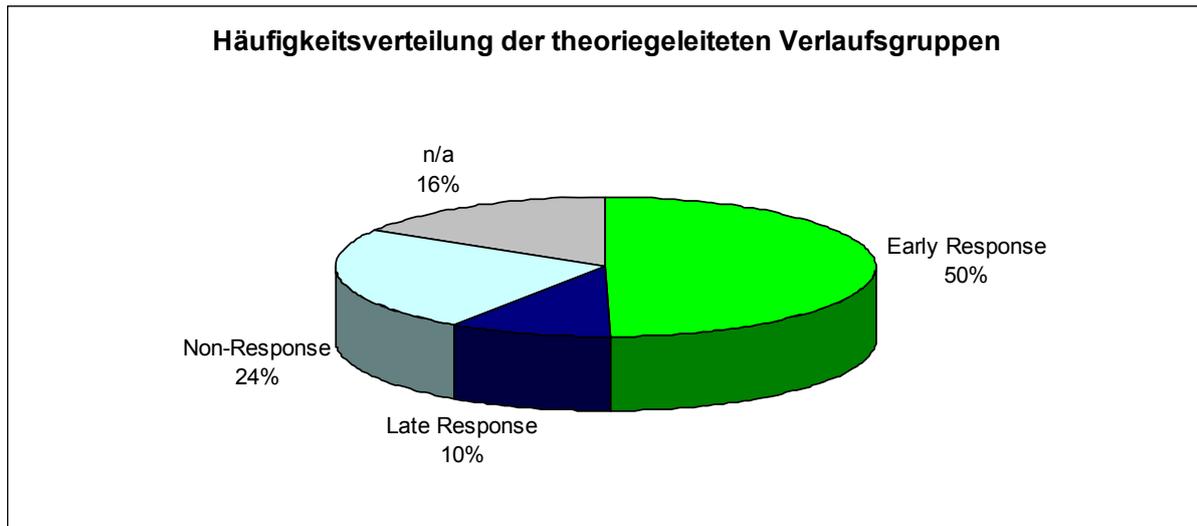


Abbildung 18: Häufigkeitsverteilung der theoriegeleiteten Verlaufsguppen

Knapp die Hälfte der Verlaufsstichprobe weist nach den theoriegeleiteten Kriterien eine Early Response auf. Das bedeutet, dass 287 Patienten nach der zweiten oder der dritten Behandlungswoche Veränderung der Symptombelastung in dem Mindestausmaß einer reliablen Verbesserung ($RCI: HEALTH-49 A_{\text{Aufnahme}} - HEALTH-49 A_{\text{Woche 2 bzw. 3}} > 0.317$) und mittleren Effektstärke (Cohens $d > 0.50$) aufweisen. Von den verbleibenden 289 Patienten erreichen noch 56 Patienten (10%) das Kriterium zum Entlasszeitpunkt und bilden die Gruppe der Late Response. Insgesamt 139 Patienten verändern sich hinsichtlich ihrer Symptombelastung weder zur zweiten bzw. dritten Woche, noch zur Entlassung in dem festgelegten Ausmaß: Die Gruppe der Non-Response umfasst 24% der Verlaufsstichprobe. Die verbleibenden 94 Patienten (16%) konnten aufgrund von fehlenden Werten in der zweiten und dritten Behandlungswoche keiner der drei inhaltlich beschreibbaren Gruppen zugeordnet werden und werden unter der Bezeichnung *n/a*, gemäß der englischen Abkürzung *not applicable* (nicht anwendbar), subsumiert.

Die *Hypothese 1* eines Mindestanteils von 5% für die a priori definierte Verlaufsguppe *Early Response* wird angenommen. Ein bedeutender Anteil von mehr als 5% der Patienten lässt

sich der theoriegeleiteten Early-Response_{TG}-Gruppe zuordnen. Die Nullhypothese einer Häufigkeitsverteilung von weniger als 5% kann mit einer 95%igen Sicherheit verworfen werden, da die Untergrenze des 95%igen Konfidenzintervalls die 5% nicht unterschreitet.

In Tabelle 19 wird die in Abschnitt 4.3.4 dargestellte Einteilung der theoriegeleiteten Verlaufsgruppenbildung wieder aufgegriffen und mit den jeweiligen Gruppengrößen dargestellt.

Tabelle 19: Schematische Darstellung der theoriegeleiteten Gruppenbildung mit den jeweiligen Gruppengrößen

		Effektstärke bei Entlassung		
		> 0.50	< 0.50	Missing
Effektstärke nach Woche 2 oder 3	> 0.50	Early Response gesamt (n = 287)		
		Early Response (n = 226)	Early Response (n = 31)	Early Response (n = 30)
	< 0.50	Late Response (n = 56)	Non-Response (n = 139)	n/a (n = 0)
	Missing ¹	n/a ² (n = 19)	n/a (n = 44)	n/a (n = 31)

¹ fehlender Wert

² nicht anwendbar: Restkategorie, für die eine Bestimmung der Responseform aufgrund fehlender Werte nicht möglich ist.

In der differenzierten Betrachtung der Verteilung der Patienten auf die einzelnen Zellen wird deutlich, dass der größte Anteil der Patienten mit einer Early Response auch zur Entlassung das hier festgelegte Kriterium der Response erfüllt (n = 226; 79% aller Patienten, die eine Early Response aufweisen). Auch von allen Patienten, die zum Zeitpunkt der Entlassung eine Verbesserung mit mindestens mittlerer Effektstärke aufweisen (n = 282), erreichten 80% der Patienten dieses Verbesserungsausmaß schon nach der 2. oder 3. Behandlungswoche. Die 94 Patienten der Gruppe n/a können alle aufgrund von fehlenden Werten in der 2. und 3. Behandlungswoche zu keiner der inhaltlichen Gruppen zugeordnet werden. Von diesen Patienten liegen 19 (18%) oberhalb und weitere 44 (41%) unterhalb des Kriteriums der Late Response. Wie schon in Abschnitt 4.3.4 erläutert, wurden diese Patienten dennoch der n/a Gruppe zugeteilt, da das Vorhandensein einer Early Response das maßgebende Kriterium für die Einteilung in eine der anderen beiden Gruppen ist. Für 31 Patienten liegen weder Daten aus der 2. oder 3. Woche, noch zur Entlassung vor.

Die hier gebildeten theoriegeleiteten Verlaufsgruppen gehen nach der beschriebenen a priori Einteilung *Early Response*, *Late Response*, *Non-Response* und *n/a* (s. Tabelle 18 und

Abbildung 18) in die Analysen der zweiten und dritten Fragestellung ein. Wie bei den explorativen Verlaufgruppen wird die Gruppe der nicht zuzuordnenden Patienten (n/a) in alle weiteren Analysen zwar eingeschlossen, die entsprechenden Ergebnisse werden jedoch nicht inhaltlich interpretiert.

Die deskriptive Darstellung der theoriegeleiteten Verlaufgruppen ist Teil der zweiten Fragestellung und wird in Abschnitt 5.6.2 dargestellt.

Zur klaren inhaltlichen und terminologischen Abgrenzung der jeweiligen theoriegeleiteten und der explorativen Verlaufgruppen werden im Folgenden die Verlaufgruppen mit TG für *theoriegeleitet* und EG für *explorativ* versehen²⁵ (bspw. Early Response_{TG} bzw. Early-Response_{TG}-Gruppe und Early Response_{EG} bzw. Early-Response_{EG}-Gruppe).

5.4 Vergleich der explorativen und der theoriegeleiteten Verlaufgruppen (Fragestellung 1 C)

In Fragestellung 1 C werden die explorativen und die theoriegeleiteten Verlaufgruppen hinsichtlich des Ausmaßes der Übereinstimmung miteinander verglichen. Tabelle 20 stellt die Kreuztabelle dar. Zur besseren Veranschaulichung sind die Zellhäufigkeiten zusätzlich in Abbildung 19 graphisch dargestellt.

²⁵ Für die Terminologie der Verlaufgruppen wurden, angelehnt an die in der Literatur verwendeten Bezeichnungen, die englischen Begriffe *Early Response*, *Delayed Response* etc. verwendet (vgl. Abschnitt 2.5.4). Die Begriffe zur Verlaufsbezeichnung werden in den meisten Studien, die ausschließlich in englischer Sprache publiziert wurden, ohne Bindestrich geschrieben. Infolgedessen wird hier auch die englische Schreibweise ohne Bindestrich gewählt. Eine Ausnahme stellt die Bezeichnung Non-Response dar, die in den meisten Fällen auch im Englischen verbunden wird. Für die vorliegende Arbeit wurde die Schreibweise der Begriffe entsprechend übernommen. Lediglich bei der Verbindung mit einem weiteren Substantiv (z.B. *Early-Response-Gruppe* oder *Early-Response-Patienten*) wird die gesamte Wortgruppe durch Bindestriche miteinander verbunden, da sich alle Wörter auf das letzte Wort beziehen und dies der besseren Lesbarkeit dient.

Tabelle 20: Kreuztabelle der Verteilung der theoriegeleiteten und explorativen Verlaufsgruppen in den jeweiligen anderen Gruppen

Explorative Verlaufsgruppen	Theoriegeleitete Verlaufsgruppen					Gesamt
	Early Response _{TG}	Late Response _{TG}	Non-Response _{TG}	n/a _{TG}		
Early Response_{EG}	<i>n</i>	50	1	0	2	53
	% von EG	94.3%	1.9%	0	3.8%	100%
	% von TG	17.4%	1.8%	0	2.1%	9.2%
	% von Gesamt	8.7%	0.2%	0	0.3%	9.2%
Delayed Response_{EG}	<i>n</i>	10	11	3	5	29
	% von EG	34.5%	37.9%	10.3%	17.2%	100%
	% von TG	3.5%	19.6%	2.2%	5.3%	5%
	% von Gesamt	1.7%	1.9%	0.5%	0.9%	5%
Non-Response_{EG}	<i>n</i>	27	1	24	9	61
	% von EG	44.3%	1.6%	39.3%	14.8%	100%
	% von TG	9.4%	1.8%	17.3%	9.6%	10.6%
	% von Gesamt	4.7%	0.2%	4.2%	1.6%	10.6%
Atypical Course_{EG}	<i>n</i>	12	2	3	5	22
	% von EG	54.5%	9.1%	13.6%	22.7%	100%
	% von TG	4.2%	3.6%	2.2%	5.3%	3.8%
	% von Gesamt	2.1%	0.3%	0.5%	0.9%	3.8%
Linear Response_{EG}	<i>n</i>	188	41	109	73	411
	% von EG	45.7%	10%	26.5%	17.8%	100%
	% von TG	65.5%	73.2%	78.4%	77.7%	71.4%
	% von Gesamt	32.6%	7.1%	18.9%	12.7%	71.4%
Gesamt	<i>n</i>	287	56	139	94	576
	% von EG	49.8%	9.7%	24.1%	16.3%	100%
	% von TG	100%	100%	100%	100%	100%
	% von Gesamt	49.8%	9.7%	24.1%	16.3%	100%

		Theoriegeleitete Verlaufsgruppen (TG)			
		Early Response _{TG}	Late Response _{TG}	Non-Response _{TG}	n/a _{TG}
Explorative Verlaufsgruppen (EG)	Early Response _{EG}				
	Late Response _{EG}				
	Non-Response _{EG}				
	Atypical Course _{EG}				
	Linear Response _{EG}				

Legende: n/a: nicht anwendbar: Restkategorie, für die eine Bestimmung der Responseform aufgrund fehlender Werte nicht möglich ist.

Abbildung 19: Graphische Darstellung der Kreuztabelle der Verteilung der theoriegeleiteten und explorativen Verlaufsgruppen in den jeweiligen anderen Gruppen

Bei der Betrachtung der Tabelle 20 und Abbildung 19 fällt auf, dass die explorative Verlaufsgruppe Linear Response_{EG} die größten Zellhäufigkeiten über alle theoriegeleiteten Verlaufsgruppen aufweist. Zwischen 66% und 78% der Patienten der theoriegeleiteten Verlaufsgruppen finden sich in dieser Gruppe wieder. Das bedeutet, dass die meisten Patienten aus den jeweiligen theoriegeleiteten Gruppen in der explorativen Gruppe liegen, die eigentlich keine Entsprechung in der theoriegeleiteten Gruppenbildung hat.

Von den 53 Patienten, die explorativ eine Early Response_{EG} aufweisen, erreichen fast alle auch das theoriegeleitete Kriterium einer Early Response_{TG} ($n = 50$; 94%). Erwartungsgemäß fällt keiner der Patienten mit einer explorativen Early Response_{EG} in die theoriegeleitete Non-Response_{TG}-Gruppe. Anders stellt sich das Bild der explorativen Non-Response_{EG} dar. Von den 61 Patienten fallen zwar 24 (40%) in die entsprechende theoriegeleitete Gruppe, 27 Patienten (44%) sind jedoch der theoriegeleiteten Early Response_{TG} zugeordnet. Das bedeutet, dass fast die Hälfte der Patienten, die über das explorative Verfahren, bei dem mehrere Messzeitpunkte einbezogen werden konnten, einer Non-Response_{EG} zugeordnet

wurden, nach der zweiten oder dritten Woche eine Veränderung mit mindestens mittlerer Effektstärke erreicht haben. Die explorativ gefundene Gruppe der Delayed Response_{EG} teilt sich mit ähnlich großen Anteilen auf die theoriegeleiteten Gruppen Late Response_{TG} (38%) und Early Response_{TG} (34%) auf.

Bezüglich der Übereinstimmung der Gruppenzuordnungen können die Zellhäufigkeiten vorab rein augenscheinlich geprüft werden. In der Abbildung 19 müssten bei hoher Übereinstimmung die diagonal verlaufenden Felder (grau unterlegt) die größten Zellhäufigkeiten und demnach große Quadrate, und alle darum herum liegenden Felder geringe bzw. keine Patientenzahlen enthalten. Um die Übereinstimmung beider Gruppierungen zu quantifizieren, wurde Cohens Kappa-Koeffizient berechnet. Der Koeffizient ist ein Maß der Übereinstimmung zweier Klassifikationen, dessen Berechnung jedoch die gleiche Anzahl an Ausprägungen bzw. Kategorien in den Gruppierungen voraussetzt. Aufgrund der fehlenden Entsprechung der Linear-Response_{EG}-Gruppe in den theoriegeleiteten Verlaufsgruppen wurden für die Berechnung von Cohens Kappa alle Patienten der Linear-Response_{EG}-Gruppe ausgeschlossen. Somit reduziert sich die Stichprobe auf $n = 165$, für die das Übereinstimmungsmaß der inhaltlich äquivalenten Gruppen berechnet werden kann. Die Ergebnisse der statistischen Prüfung auf Übereinstimmung sind in Tabelle 21 dargestellt.

Tabelle 21: Zusammenhang zwischen den explorativen und den theoriegeleiteten Verlaufsgruppen

Gestestete Variablen	<i>n</i> gesamt	Chi ² (<i>df</i>)	Signifikanz <i>p</i>	Cohens Kappa
Explorative * theoriegeleitete Verlaufsgruppen	165	81.095 (9)	< .001***	0.357

df: Freiheitsgrade (degrees of freedom); *p*: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Die hier erzielte Übereinstimmung der beiden Klassifizierungen mit einem Kappa von 0.36 ist nach dem von Landis und Koch (Landis & Koch, 1977) formulierten Kriterium als ausreichend zu bewerten (Kappa-Koeffizienten: 0.21 - 0.40). Die von andere Autoren vorgeschlagene Interpretationsgrenze einer ausreichend vorhandenen Übereinstimmung bei einem Kappa von 0.40 (Cicchetti, 1994; zitiert nach Wirtz & Kutschmann, 2007) wurde jedoch nicht erreicht.

5.5 Wöchentliche Verbesserungsdaten (Fragestellung 1 D)

Neben der explorativen und theoriegeleiteten Verlaufsgruppenbildung stellt sich die Frage danach, wie viele Patienten sich *in welchem zeitlichen Verlauf* anhand von *unterschiedlichen Kriterien verbessern*. Zur Beantwortung werden die wöchentlichen Verbesserungsdaten für

die drei Kriterien (1) Verbesserung mit einer mittleren Effektstärke, (2) Verbesserung um 50% und (3) Verbesserung mit klinischer Signifikanz jeweils in einer Tabelle dargestellt. Tabelle 22 bis Tabelle 24 enthalten jeweils die Anteile der Patienten, die mit jeder Behandlungswoche erstmals das Kriterium erreichen, sowie die mittlere Behandlungsdauer der jeweiligen Patientengruppe. Zusätzlich wird die Anzahl der Patienten angegeben, die das Kriterium im gesamten Behandlungszeitraum nicht erfüllt, ebenso wie die Behandlungsdauer (in drei Kategorien) dieser Patientengruppe. In Tabelle 22 zeigt sich, dass annähernd ein Drittel der Verlaufsstichprobe (32%) nach der ersten Behandlungswoche eine Verbesserung mit mindestens mittlerer Effektstärke aufweist. Nach der dritten Woche hat über die Hälfte der Patienten (55%) mindestens zu einem der drei Messzeitpunkte das Kriterium erreicht. Bis zum Ende der Behandlung haben über zwei Drittel (68%) der Patienten das Kriterium mindestens einmal erreicht, die übrigen 33% der Patienten werden zum größten Teil nach 5 bis 6 Wochen entlassen. Zur Prüfung, wie viele Patienten das Kriterium zum Entlasszeitpunkt erreicht haben, unabhängig von einem vorherigen Erreichen oder Nichterreichen, wurde die absolute Häufigkeit der Patienten bei Entlassung dargestellt. Insgesamt 52% der Patienten weisen im Vergleich der zwei Zeitpunkte Aufnahme und Entlassung eine Verbesserung von mindestens mittlerer Effektstärke auf.

Tabelle 22: Verteilung bis zum Erreichen des Verbesserungskriteriums: *Verbesserung mit einer mittleren Effektstärke* ($n = 576$)

Woche, in der das Kriterium erstmals erreicht wird:	Anzahl	Prozent der Gesamtstichprobe	Prozent kumuliert	Behandlungsdauer in Tagen (M (SD))
1 ($n = 548$)	182	31.6	31.6	42.1 (8.6)
2 ($n = 536$)	94	16.3	47.9	42.8 (7.8)
3 ($n = 468$)	39	6.8	54.7	42.5 (7.8)
4 ($n = 399$)	36	6.3	60.9	42.8 (6.9)
5 ($n = 202$)	25	4.3	65.3	45.5 (6.0)
6 ($n = 71$)	7	1.2	66.5	52.6 (4.8)
7 ($n = 20$)	6	1.0	67.5	52.8 (3.7)
8 ($n = 14$)	0	--	67.5	--
9 ($n = 5$)	0	--	67.5	--
bei Entlassung erreicht	301	52.3	--	--
innerhalb von neun Wochen <i>kein Mal</i> erreicht	187	32.5	100.0	40.0 (9.0)
davon entlassen nach:	Anzahl	Prozent von „nicht erreicht“	Prozent der Gesamtstichprobe	
3-4 Wochen	35	18.7	6.1	
5-6 Wochen	120	64.2	20.8	
7-9 Wochen	32	17.1	5.6	

M: Mittelwert; SD: Standardabweichung

Betrachtet man die Verbesserungsraten um 50% der Ausgangsbelastung (Tabelle 23), wird deutlich, dass 17% der Patienten nach der ersten Woche eine Belastungsreduktion um die Hälfte erreichen. Nach der 4. (49%) bis 5. Behandlungswoche (55%) hat die Hälfte der Patienten mindestens einmal das Kriterium der Verbesserung um 50% erreicht. 40% der Patienten erreicht das Kriterium während der gesamten Behandlungszeit nicht. Alleine zum Entlasszeitpunkt, d.h. unabhängig vom vorherigen Verlauf, erreichen 274 (48%) Patienten das Kriterium einer 50%igen Verbesserung.

Tabelle 23: Verteilung bis zum Erreichen des Verbesserungskriteriums: *Verbesserung um 50 Prozent der Ausgangsbelastung (n = 576)*

Woche, in der das Kriterium erstmals erreicht wird:	Anzahl	Prozent der Gesamtstichprobe	Prozent kumuliert	Behandlungsdauer in Tagen (M/SD)
1 (n = 548)	97	16.8	16.8	40.9 (9.0)
2 (n = 536)	78	13.5	30.4	39.6 (8.6)
3 (n = 468)	58	10.1	40.5	43.7 (7.6)
4 (n = 399)	47	8.2	48.6	40.6 (6.9)
5 (n = 202)	38	6.6	55.2	44.5 (6.0)
6 (n = 71)	10	1.7	56.9	47.2 (6.7)
7 (n = 20)	10	1.7	58.7	52.8 (3.7)
8 (n = 14)	2	0.3	59.0	57.5 (3.5)
9 (n = 5)	1	0.2	59.2	24.0
bei Entlassung erreicht	274	47.6	--	--
innerhalb von neun Wochen <i>kein Mal</i> erreicht	235	40.8	100.0	41.9 (8.6)
davon entlassen nach:	Anzahl	Prozent von „nicht erreicht“	Prozent der Gesamtstichprobe	
3-4 Wochen	33	14.0	5.7	
5-6 Wochen	147	62.6	25.5	
7-9 Wochen	55	23.4	9.5	

M: Mittelwert; SD: Standardabweichung

Das Kriterium einer klinisch signifikanten Verbesserung stellt das strengste der drei Kriterien dar. Da das Kriterium das Unterschreiten des Cut-Off-Wertes (HEALTH-49 Modul A: Cut-Off = 0.685; Rabung et al., 2008; vgl. auch 4.3.1 und 4.5.4) beinhaltet, können die Werte der klinischen Signifikanz nur für die Patienten ermittelt werden, die bei Behandlungsbeginn oberhalb dieses Wertes liegen. Aus diesem Grund reduziert sich die Stichprobe um 19% auf $n = 464$. Nach der ersten Woche erreichen 67 Patienten (14%) eine Veränderung mit dem Ausmaß einer reliablen Verbesserung und liegen unterhalb des Cut-Off-Wertes. Nach der dritten Woche haben mehr als ein Drittel der Patienten (36%) das Kriterium mindestens einmal erfüllt, nach der vierten Woche 44%. Im Verlauf der folgenden Behandlungswochen zeigt sich, dass jeweils nur noch ein geringer Anteil an Patienten dieses Kriterium *erstmalig*

erreicht. Nach der siebten Woche haben 50% der Patienten das Kriterium mindestens einmal erreicht, die andere Hälfte der Patienten erreicht dieses Kriterium während des gesamten Behandlungszeitraums nicht. Auch hier zeigt sich bei alleiniger Betrachtung des Entlasszeitpunkts, dass 41% der Patienten bei Behandlungsende eine klinisch signifikante Verbesserung aufweist (Tabelle 24).

Tabelle 24: Verteilung bis zum Erreichen des Verbesserungskriteriums: *Klinisch signifikante Verbesserung* ($n = 464$)

Woche, in der das Kriterium erstmals erreicht wird:	Anzahl	Prozent der Gesamtstichprobe	Prozent kumuliert	Behandlungsdauer in Tagen (M / SD)
1 ($n = 440$)	67	14.4	14.4	40.8 (8.2)
2 ($n = 432$)	56	12.1	26.5	39.7 (8.8)
3 ($n = 379$)	45	9.7	36.2	43.6 (7.5)
4 ($n = 324$)	37	8.0	44.2	40.9 (8.2)
5 ($n = 172$)	30	6.5	50.6	44.5 (6.8)
6 ($n = 61$)	9	1.9	52.6	47.4 (5.7)
7 ($n = 18$)	3	0.6	53.2	51.3 (8.1)
8 ($n = 13$)	0	--	53.2	--
9 ($n = 5$)	0	--	53.2	--
bei Entlassung erreicht	190	40.9	--	--
innerhalb von neun Wochen <i>kein Mal</i> erreicht	217	46.8	100.0	42.9 (8.6)
davon entlassen nach:	Anzahl	Prozent von „nicht erreicht“	Prozent der Gesamtstichprobe	
3-4 Wochen	25	11.5	5.4	
5-6 Wochen	130	59.9	28.0	
7-9 Wochen	62	28.6	13.4	

M: Mittelwert; SD: Standardabweichung

Die jeweiligen kumulierten Verbesserungsdaten im zeitlichen Verlauf sind in Abbildung 20 dargestellt.

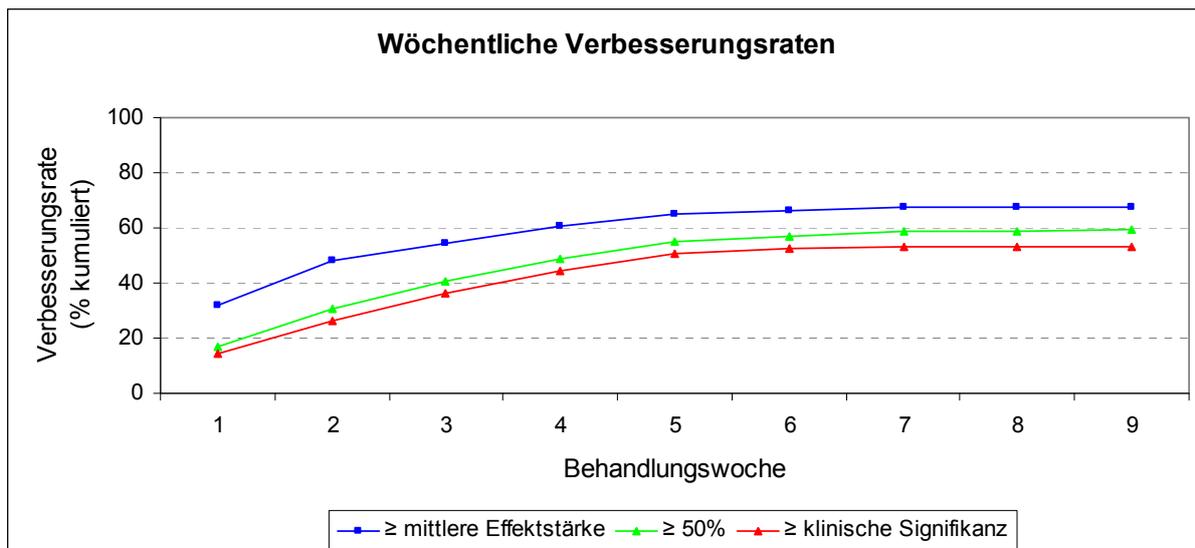


Abbildung 20: Wöchentliche Verbesserungsrate nach den Kriterien (1) Verbesserung mit mindestens mittlerer Effektstärke; (2) Verbesserung um mindestens 50 % der Ausgangsbelastung; (3) Verbesserung mit mindestens klinischer Signifikanz

In der Verlaufstichprobe weisen insgesamt 31 Patienten (5.4%) zum Ende der Behandlung eine Verschlechterung der Symptombelastung auf. Das bedeutet, dass sich diese Patienten von der Aufnahme bis zur Entlassung in dem Mindestausmaß des RCI²⁶ verschlechtert haben ($\text{HEALTH-49 } A_{\text{Aufnahme}} - \text{HEALTH-49 } A_{\text{Entlassung}} < -0.317$). Von diesen Patienten befinden sich 30 zur Entlassung in einem klinisch auffälligen Bereich (oberhalb des Cut-Off-Wertes von 0.685). Hinsichtlich der explorativen Verlaufgruppen verteilen sich diese Patienten auf die Gruppen Non-Response_{EG} ($n = 17$; 55%), Linear Response_{EG} ($n = 9$; 29%), Atypical Course ($n = 3$; 10%) und Delayed Response_{EG} ($n = 2$; 6%). In den theoriegeleiteten Verlaufgruppen sind sie in den Gruppen Non-Response_{TG} ($n = 25$; 81%), n/a ($n = 4$; 12%) und Early Response_{TG} ($n = 2$; 6%) wiederzufinden. Aufgrund der relativ geringen Anzahl an verschlechterten Patienten ($n < 50$), werden keine gesonderten Analysen für diese Patienten durchgeführt (vgl. Abschnitt 4.3.4).

5.6 Vorhersage der Verlaufgruppen (Fragestellung 2)

Zur Untersuchung der Frage nach der Vorhersage der Zugehörigkeit zu den Verlaufgruppen (Fragestellung 2), wurden zunächst univariate Analysen zur Unterschiedstestung durchgeführt (5.6.1 und 5.6.2). Dabei wurde untersucht, ob sich die fünf explorativen bzw. vier theoriegeleiteten Verlaufgruppen in soziodemographischen und krankheitsbezogenen Merkmalen signifikant voneinander unterscheiden. Zur Signifikanztestung bei kontinuierlichen Variablen wurden Varianzanalysen und das entsprechende Effektmaß Eta^2

²⁶ Vgl. auch Abschnitt 2.3.1 und 4.5.4.

berechnet. Für kategoriale Variablen kamen Chi²-Tests mit der Effektgröße Phi und Fishers exakte Tests zum Einsatz.

In einem zweiten Schritt wurden die postulierten *Hypothesen 2* und *3* zur Vorhersage der Zugehörigkeit zu den jeweiligen Verlaufsgruppen über inhaltlich und empirisch fundierte Patientenmerkmale überprüft (5.6.3 und 5.6.4).

5.6.1 Vergleiche der explorativen Verlaufsgruppen

Die Ergebnisse der Merkmalsvergleiche zwischen den explorativen Verlaufsgruppen sind im Anhang in Tabelle C.1 bis Tabelle C.7 dargestellt.

5.6.1.1 Soziodemographische Merkmale

Hinsichtlich der soziodemographischen Merkmale bestehen keine signifikanten Unterschiede zwischen den Verlaufsgruppen in den Variablen Alter, Familienstand und monatliches Nettohaushaltseinkommen. Der Anteil an Frauen ist in den beiden Gruppen Early und Non-Response_{EG} mit 91% und 82% am höchsten, in den Gruppen Delayed und Linear Response_{EG} hingegen liegt der Frauenanteil bei 72% und 75%. Der Unterschied erreicht jedoch nicht das Niveau einer statistischen Signifikanz ($p = .092$, $\phi = -.106$). Deutliche Unterschiede zwischen den explorativen Verlaufsgruppen lassen sich in den Variablen Schulabschluss, Berufsabschluss und Erwerbstätigkeit finden. Die Patienten mit verzögerten oder ausbleibendem Ansprechen auf die Behandlung (Delayed Response_{EG} und Non-Response_{EG}) haben mit 38% und 43% deutlich häufiger einen Hauptschulabschluss als höchsten Schulabschluss als die Patienten aus den Gruppen Early- und Linear Response_{EG} (22% und 20%). Entsprechend befinden sich mehr Patienten mit höherem Schulabschluss (Realschule/Polytechnische Oberschule oder Abitur/Fachabitur) in den beiden letztgenannten (Early Response_{EG} und Linear Response_{EG} jeweils 77%) als in den Gruppen Delayed Response_{EG} (57%) und Non-Response_{EG} (54%). Patienten mit Abitur oder Fachabitur sind anteilig am wenigsten in der Delayed-Response_{EG}-Gruppe (11%) und am stärksten in der Linear-Response_{EG}-Gruppe (37%) vertreten. Die Gruppenunterschiede sind hoch signifikant mit annähernd mittlerer Effektstärke [$\chi^2 (12, n = 562) = 39.848$; $p < .001^{***}$, $\phi = .267$]. Die meisten Patienten der Gesamtstichprobe geben als höchsten Berufsabschluss eine abgeschlossene Lehre an, was sich auch in den einzelnen Verlaufsgruppen widerspiegelt. Patienten mit einem Hochschulabschluss sind allerdings mit einem Anteil von 23% weit häufiger in der Gruppe der Linear Response_{EG} zu finden als in den anderen drei Gruppen (5% - 12%). Dagegen sind in der Non-Response_{EG}-Gruppe, ähnlich wie in der Delayed-Response_{EG}-Gruppe, deutlich mehr Patienten ohne Berufsabschluss (17% und 12%) als in den Gruppen Early Response_{EG} (8%) und Linear Response_{EG} (6%) [$\chi^2 (20, n =$

531) = 34.999, $p = .008^{**}$, $\phi = .255$]. Ein ähnliches Bild zeigt sich für die Erwerbstätigkeit: In der Non-Response_{EG}-Gruppe und in der Delayed-Response_{EG}-Gruppe sind weniger Patienten Vollzeit oder Teilzeit berufstätig (55% und 65%) als in den beiden anderen Verlaufgruppen (Linear Response_{EG}: 82%; Early Response_{EG}: 79%). Entsprechend ist der Anteil an nicht erwerbstätigen Patienten (arbeitslos, in Rente, in Ausbildung, Hausfrau/-mann) in den beiden erstgenannten Verlaufgruppen höher [$\chi^2 (8, n = 521) = 26.151, p = .001^{**}, \phi = .230^+$].

5.6.1.2 Sozialmedizinische Merkmale

In den sozialmedizinischen Merkmalen sind bedeutende Unterschiede zwischen den Verlaufgruppen zu finden. Die Patienten der Non-Response_{EG}- und der Delayed-Response_{EG}-Gruppe weisen mit durchschnittlich 15 Wochen deutlich höhere Arbeitsunfähigkeitszeiten in den 6 Monaten vor Behandlungsbeginn als die Patienten der Linear-Response_{EG}-Gruppe ($M = 8; SD = 9$) und der Early-Response_{EG}-Gruppe ($M = 10; SD = 10$) auf [$F(4,533) = 10.366, p < .001^{***}, \eta^2 = .072$]. In diesen beiden Gruppen ist der Anteil der Patienten, die wegen der aktuellen Erkrankung schon eine oder mehrere stationäre psychotherapeutische Behandlungen durchgeführt haben, deutlich niedriger (Linear Response_{EG}: 8%, Early Response_{EG}: 8%) als in den beiden erstgenannten Verlaufgruppen (Non-Response_{EG}: 18%, Delayed Response_{EG}: 21%) [$\chi^2 (4, N = 574) = 9.610, p = .036^*, \phi = .132$]. Dahingegen bestehen keine signifikanten Unterschiede in den Häufigkeiten ambulanter psychotherapeutischer Vorbehandlungen zwischen den Verlaufgruppen.

5.6.1.3 Diagnosen

Hinsichtlich der psychischen Diagnosen bestehen für fast alle Indikationsgruppen signifikante Unterschiede zwischen den explorativen Verlaufgruppen (s. Anhang: Tabelle C.4). Die Diagnosegruppen wurden nicht ausschließlich aus den Erstdiagnosen sondern aus allen vergebenen Diagnosestellen gebildet (vgl. Abschnitt 4.3.3.1). Die Verteilung der explorativen Verlaufgruppen in den Diagnosegruppen der Hauptindikationen ist in Abbildung 21 dargestellt.

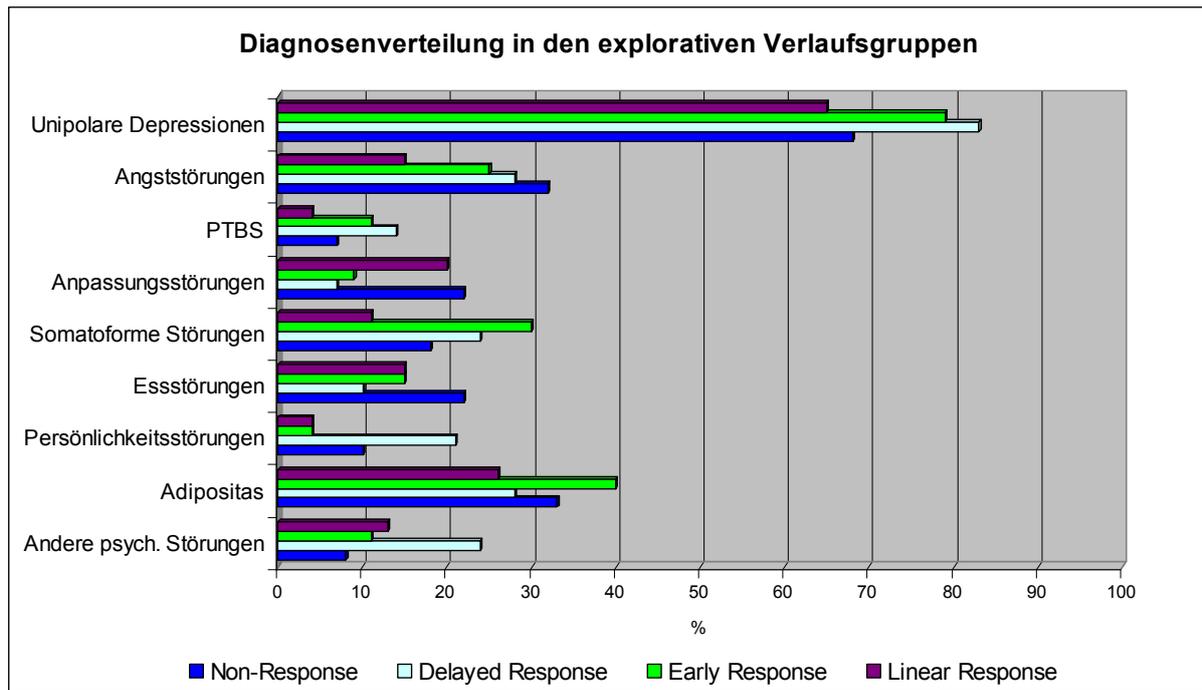


Abbildung 21: Verteilung der explorativen Verlaufsgruppen in den Diagnosegruppen

Die unipolaren Depressionen sind mit 83% in der Gruppe der Delayed Response_{EG} und mit 79% in der Gruppe der Early Response_{EG} stärker vertreten als in der Linear-Response_{EG}-Gruppe (65%) und der Non-Response_{EG}-Gruppe (68%). Eine ähnliche Verteilung zwischen den vier Verlaufsgruppen zeigt sich für die Häufigkeiten von Posttraumatischen Belastungsstörungen und Somatoformen Störungen. Patienten mit Angststörungen sind am wenigsten in der Linear-Response_{EG}-Gruppe (15%) zu finden (Anteile in den anderen Gruppen 25% - 32%). Anpassungsstörungen kommen doppelt so häufig in der Non-Response_{EG}-Gruppe (22%) und der Linear-Response_{EG}-Gruppe (20%) wie in den Gruppen Early Response_{EG} (9%) und Delayed Response_{EG} (7%) vor. Patienten mit somatoformen Störungen sind am stärksten in der Early-Response_{EG}-Gruppe vertreten (30%) und am geringsten in der Linear-Response_{EG}-Gruppe (14%). In der Delayed-Response_{EG}-Gruppe sind weniger Patienten mit einer Essstörung (10%) als in den anderen drei Verlaufsgruppen (15% - 22%) zu finden, wohingegen die Persönlichkeitsstörungen in dieser Verlaufsgruppe am häufigsten vorkommen (21% vs. 4% - 10%).

Bei der Diagnoseverteilung muss berücksichtigt werden, dass hier alle vergebenen Diagnosen aufgeführt sind, sodass aus den beschriebenen Ergebnissen Aussagen zur Komorbidität abgeleitet werden können. In Abbildung 21 ist zu sehen, dass die Gruppe der Linear Response_{EG} über alle Diagnosen geringer besetzt ist als die anderen drei Verlaufsgruppen. Dies spiegelt sich auch in den Ergebnissen zu der Anzahl vergebenen F-Diagnosen in den Verlaufsgruppen wider: Während in der Linear-Response_{EG}-Gruppe 47% der Patienten eine psychische Komorbidität (mehr als eine F-Diagnose) aufweisen, sind die

anderen drei Verlaufgruppen durch einen weit höheren Anteil an psychisch komorbiden Patienten gekennzeichnet (Early Response_{EG}: 77%, Delayed Response_{EG}: 79%, Non-Response_{EG}: 80%) [$\chi^2(4, N = 571) = 43.033, p < .001^{***}, \phi = .275$].

Insgesamt lässt sich festhalten, dass die Diagnosegruppen in den explorativen Verlaufgruppen zwar in signifikant unterschiedlichen Häufigkeiten vorkommen, unter Berücksichtigung der Mehrfachdiagnosen kann jedoch für keine Diagnose eine klare Zuordnung zu einer Verlaufgruppe vorgenommen werden. Die Verteilung der Diagnosen, gruppiert nach den explorativen Verlaufgruppen, ist in Abbildung 22 dargestellt.

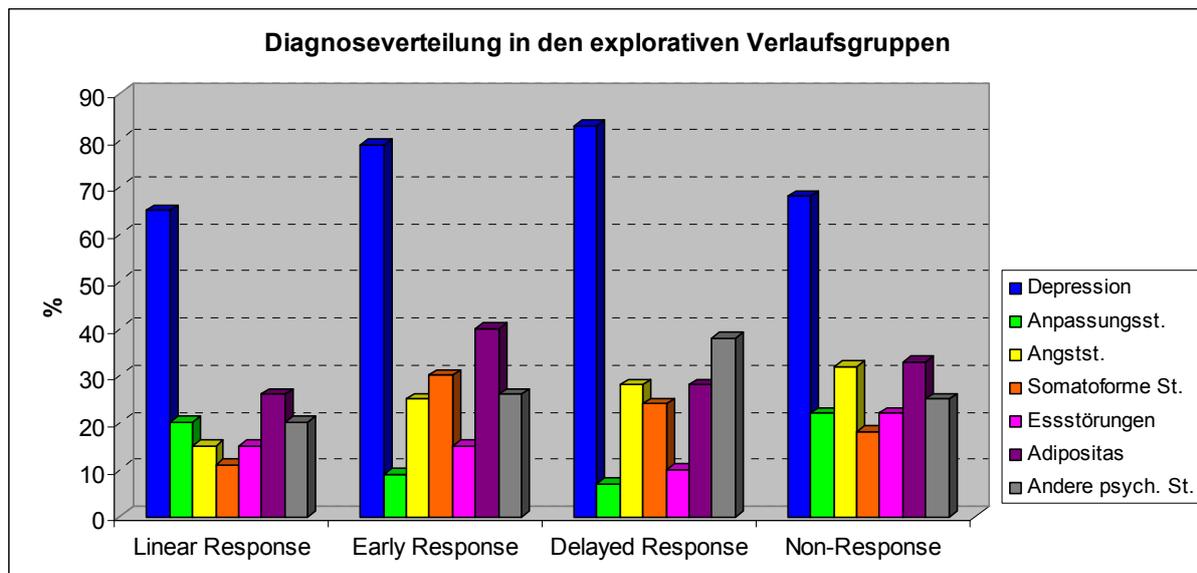


Abbildung 22: Verteilung der Diagnosegruppen in den explorativen Verlaufgruppen

5.6.1.4 Krankheitsbezogene Merkmale

Zur Untersuchung von Unterschieden in krankheitsbezogenen Merkmalen zwischen den Verlaufgruppen wurden psychometrische Skalenwerte der Symptombelastung (HEALTH-49 A), der psychischen Aspekte der gesundheitsbezogenen Lebensqualität (SF-8 Psych.), der Psychotherapiemotivation (FPTM-23) und der allgemeinen Selbstwirksamkeitserwartung (SWE) zu Beginn der Behandlung analysiert (s. Anhang: Tabelle C.5).

Da der Ausgangswert im HEALTH-49 A einen Parameter zu Identifikation der explorativen Verlaufgruppen darstellt (vgl. Abschnitt 4.5.2), erweisen sich die Unterschiede zwischen den explorativen Verlaufgruppen in der Symptombelastung zu Behandlungsbeginn als hoch signifikant [$F(4,571) = 139.573, p < .001^{***}, \eta^2 = .494^{+++}$], mit dem höchsten Skalenmittelwert in der Delayed-Response_{EG}-Gruppe ($M = 2.8; SD = 0.6$), gefolgt von der Early-Response_{EG}-Gruppe ($M = 2.3; SD = 0.5$) und der Non-Response_{EG}-Gruppe ($M = 2.0; SD = 0.6$) und der geringsten mittleren Ausprägung der Symptombelastung in der Linear-Response_{EG}-Gruppe ($M = 1.1; SD = 0.5$). Von den 112 Patienten (19.4% der Verlaufstichprobe), die bei Aufnahme

im HEALTH-49 A unterhalb des Cut-Off-Wertes von 0.685 lagen, wurden nahezu alle Patienten ($n = 110$) der Gruppe Linear Response_{EG} zugeordnet. Diese entsprechen 27% der Gesamtgruppe der 411 Patienten mit einer Linear Response_{EG}. Die verbleibenden zwei Patienten sind der Gruppe Atypical Course zugeordnet.

Für die psychische Summenskala des SF-8 zeigt sich ein ähnliches Muster: Patienten mit einer Linear Response_{EG} weisen zu Beginn der Behandlung die höchsten Werte der psychischen Aspekte gesundheitsbezogener Lebensqualität auf ($M = 35$; $SD = 12$), Patienten aus den Gruppen Early Response_{EG} ($M = 24$; $SD = 8$) und Delayed Response_{EG} ($M = 23$; $SD = 7$) die niedrigsten [$F(4,560) = 24.275$, $p < .001^{***}$, $\eta^2 = .148^{+++}$].

Ebenso sind die Werte der allgemeinen Selbstwirksamkeitserwartung in der Linear-Response_{EG}-Gruppe am höchsten ($M = 25$; $SD = 6$), gefolgt von den Gruppen Non-Response_{EG} ($M = 21$; $SD = 6$), Early Response_{EG} ($M = 20$; $SD = 5$) und Delayed Response_{EG} ($M = 19$; $SD = 5$) [$F(4,567) = 16.034$, $p < .001^{***}$, $\eta^2 = .102^{++}$].

Die Psychotherapiemotivation wurde zu Behandlungsbeginn mit sechs Skalen des FPTM-23 erhoben. Die Ergebnisse zeigen für die vier Skalen *psychischer Leidensdruck*, *Hoffnung*, *Verneinung psychischer Hilfsbedürftigkeit* und *symptombezogene Zuwendung durch Andere* signifikante Unterschiede zwischen den explorativen Verlaufgruppen. Die Patienten der Linear-Response_{EG}-Gruppe zeigen einen deutlich niedrigeren Leidensdruck [$F(4,567) = 21.983$, $p < .001^{***}$, $\eta^2 = .134^{++}$], eine geringere Verneinung psychischer Hilfsbedürftigkeit [$F(4,567) = 3.856$, $p = .004^{**}$, $\eta^2 = .026^+$], eine höhere Hoffnung [$F(4,564) = 9.846$, $p < .001^{***}$, $\eta^2 = .065^{++}$] und symptombezogene Zuwendung [$F(4,564) = 3.789$, $p = .005^{**}$, $\eta^2 = .026^+$] als die Patienten der anderen drei Verlaufgruppen. Der niedrige Leidensdruck und die hohe symptombezogene Zuwendung sprechen für eine geringe Psychotherapiemotivation, während eine geringe Verneinung und hohe Werte in der Hoffnung für eine hohe Psychotherapiemotivation stehen. In den Skalen *Wissen* und *Initiative* des FPTM-23 lassen sich keine signifikanten Gruppenunterschiede finden.

Insgesamt zeigt sich, dass die Patienten mit einer Linear Response_{EG} im Vergleich zu den anderen drei Gruppen zu Behandlungsbeginn die geringste Symptombelastung, die höchste Lebensqualität, die höchste Selbstwirksamkeitserwartung und, abgesehen von dem geringen Leidensdruck, eine stärker ausgeprägten Psychotherapiemotivation aufweisen.

5.6.1.5 Prozessmerkmale

Im Folgenden werden die Prozessmerkmale Behandlungsdauer, Behandlungsdosis- und art, sowie die therapeutische Beziehung in der ersten Phase der Behandlung für die vier Verlaufgruppen gegenübergestellt (s. Anhang: Tabelle C.3 und Tabelle C.6).

Die mittlere Behandlungsdauer beträgt für die gesamte Verlaufsstichprobe 41.9 Tage ($SD = 9$). Die längsten Behandlungszeiten mit durchschnittlich 44.7 Tagen ($SD = 9$) finden sich bei der Gruppe der Non-Response_{EG}, gefolgt von der Delayed-Response_{EG}-Gruppe mit durchschnittlich 43.9 Tagen ($SD = 8$). Dahingegen werden die Patienten mit einer Early Response_{EG} ($M = 41.7$; $SD = 9$) und einer Linear Response_{EG} ($M = 41.6$; $SD = 9$) etwas kürzer behandelt. Die Gruppenunterschiede sind signifikant mit kleiner Effektstärke [$F(4,571) = 2.827$, $p = .024^*$, $\eta^2 = .019$]. Im Hinblick auf die Dosis psychotherapienaher Maßnahmen, erfasst über die Gesamtdauer aller während des Aufenthalts erhaltenen psychotherapienahen Behandlungen (vgl. Abschnitt 4.3.3.3), finden sich ebenfalls signifikante Gruppenunterschiede mit kleiner Effektstärke [$F(4,519) = 2.415$, $p = .048^*$, $\eta^2 = .019$]: Hier sind es jedoch die Patienten der Early-Response_{EG}-Gruppe, die mit durchschnittlich insgesamt 38.6 Stunden die höchste Dosis psychotherapienaher Maßnahmen erhalten haben (Gesamtdauer in Minuten: $M = 2\,313$; $SD = 873$). Gefolgt wird diese von der Gruppe der Delayed Response_{EG} mit durchschnittlich insgesamt 37.9 Stunden psychotherapienahen Maßnahmen ($M = 2\,275$; $SD = 1\,083$). Die Non-Response_{EG}-Gruppe, die die längsten Behandlungsdauern aufweist, erhält während der gesamten Behandlung hingegen eine Dosis von nur 35.8 Stunden ($M = 2\,149$; $SD = 975$). Die geringste Dosis psychotherapienaher Maßnahmen ist mit durchschnittlich 33.5 Stunden bei der Linear-Response_{EG}-Gruppe zu verzeichnen ($M = 2\,008$; $SD = 791$). Geht man davon aus, dass sich die Maßnahmen regelmäßig auf die Behandlungswochen verteilen, so kann in der Early-Response_{EG}-Gruppe von einer durchschnittlichen Dosis von 6.5 Stunden pro Woche²⁷, hingegen in der Non-Response_{EG}-Gruppe von durchschnittlich 5.6 Stunden pro Woche ausgegangen werden. Für die Patienten mit einer Linear Response_{EG} wären es entsprechend 5.6 Stunden pro Woche und bei der Delayed Response_{EG} 6 Stunden pro Woche.

Um ein differenziertes Bild über psychotherapeutische Behandlungen in Hinblick auf die therapeutische Ausrichtung zu erhalten, wurde die Art der psychotherapienahen Behandlungsdosis über das Verhältnis von verhaltenstherapeutischen und psychodynamisch orientierten Therapiemaßnahmen erfasst. Die Delayed-Response_{EG}-Gruppe hat mit einem mittleren Anteil von 27% an verhaltenstherapeutischen Maßnahmen den geringsten Anteil und somit auch den höchsten Anteil an psychodynamischen Behandlungsmaßnahmen (73%) erhalten. In den anderen Gruppen liegt der Anteil erhaltener verhaltenstherapeutischer Maßnahmen zwischen 31% und 34%. Insgesamt sind die Gruppenunterschiede jedoch nicht signifikant [$F(4,508) = 0.411$, $p = .801$, $\eta^2 = .003$].

²⁷ Um die Stundenzahl pro Woche zu ermitteln wurde die Gesamtstundenanzahl der Verlaufsstichprobe durch die mittlere Behandlungsdauer (in Tagen) der jeweiligen Gruppe dividiert und anschließend mit 7 multipliziert.

Als personenbezogenes Prozessmerkmal wurde die therapeutische Beziehung nach der zweiten und dritten Behandlungswoche mit dem HAQ erhoben. Hier zeigt sich erwartungsgemäß, dass die Güte der therapeutischen Beziehung von den Patienten der Gruppe der Early Response_{EG} ($M = 29.7$; $SD = 5$) und der Gruppe der Linear Response_{EG} ($M = 29.6$; $SD = 5$) höher bewertet wird als in der Gruppe der Non-Response_{EG} ($M = 27.0$; $SD = 6$) und der Delayed Response_{EG} ($M = 26.6$; $SD = 5$). Die Gruppen unterscheiden sich signifikant mit kleiner Effektstärke [$F(4,547) = 6.114$, $p < .001^{***}$, $\eta^2 = .043^+$].

5.6.2 Vergleiche der theoriegeleiteten Verlaufgruppen

Wie im vorangegangenen Abschnitt 5.6.1 wurden alle Patienten- und Prozessmerkmale in Hinblick auf Unterschiede zwischen den theoriegeleiteten Verlaufgruppen untersucht. Die Ergebnistabellen sind im Anhang in Tabelle D.1 bis Tabelle D.7 zu finden.

5.6.2.1 Soziodemographische Merkmale

In den soziodemographischen Merkmalen zeigt sich lediglich bei der Erwerbstätigkeit in signifikanter Unterschied zwischen den theoriegeleiteten Verlaufgruppen: In den Gruppen Late Response_{TG} (28%) und Early Response_{TG} (25%) ist der Anteil der nicht erwerbstätigen Patienten höher als in der Non-Response_{TG}-Gruppe (15%) [$\chi^2(6, n = 521) = 13.344$, $p = .038^*$, $\phi = .160^+$].

In den Merkmalen Alter, Geschlecht, Familienstand und Haushaltseinkommen bestehen keine signifikanten Unterschiede zwischen den theoriegeleiteten Verlaufgruppen.

5.6.2.2 Sozialmedizinische Merkmale

In den sozialmedizinischen Angaben zu Arbeitsunfähigkeitszeiten sowie stationären und ambulanten Vorbehandlungen aufgrund der psychischen Erkrankung, die Anlass für die untersuchte stationäre Behandlung war, bestehen keine signifikanten Unterschiede zwischen den theoriegeleiteten Verlaufgruppen.

5.6.2.3 Diagnosen

Die Ergebnisse zur Verteilung der theoriegeleiteten Verlaufgruppen auf alle vergebenen psychischen Diagnosen sind in Abbildung 23 sowie im Anhang in Tabelle D.4 dargestellt. Signifikante Gruppenunterschiede in den Häufigkeiten finden sich bei den Diagnosen unipolare Depressionen und Anpassungsstörungen: Während Patienten mit einer unipolaren Depression häufiger in der Late-Response_{TG}- (79%) und der Early-Response_{TG}-Gruppe (74%) als in der Non-Response_{TG}-Gruppe (57%) vertreten sind [$\chi^2(3, n = 571) = 19.757$, $p < .001^{***}$, $\phi = .186^+$], verhält es sich bei den Anpassungsstörungen mit 25% in der Non-

Response_{TG}-Gruppe und 14% und 11% in der Early-Response_{TG}- und Late-Response_{TG}-Gruppe genau umgekehrt [$\chi^2(3, n = 571) = 10.816, p = .013^*, \phi = .138^+$]. Alle anderen Diagnosegruppen unterscheiden sich nicht signifikant hinsichtlich der Häufigkeiten in den theoriegeleiteten Verlaufsgruppen.

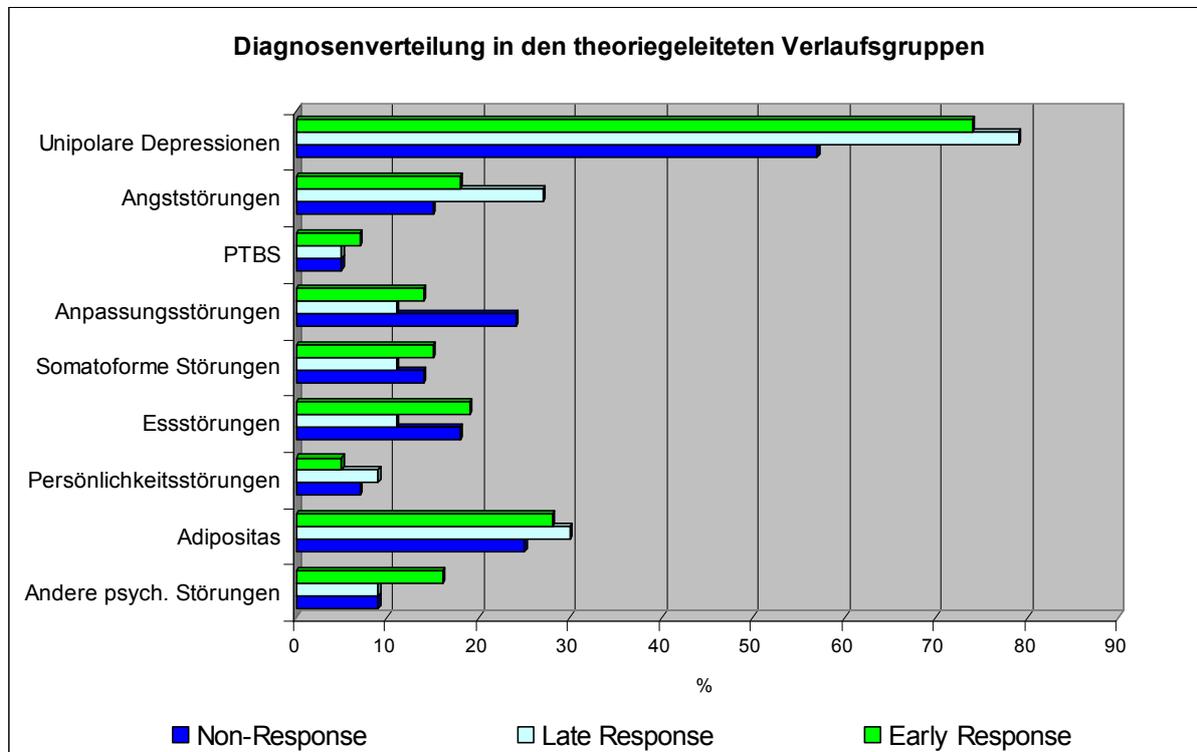


Abbildung 23: Verteilung der theoriegeleiteten Verlaufsgruppen in den Diagnosegruppen

Patienten mit psychischer Komorbidität sind mit 60% und 59% stärker in den Gruppen Early Response_{TG} und Late Response_{TG} als in der Non-Response_{TG}-Gruppe (51%) vertreten [$\chi^2(3, n = 571) = 7.869, p = .049^*, \phi = .117^+$].

Insgesamt zeigt sich auch hier, ähnlich wie bei den explorativen Verlaufsgruppen, dass zwar bedeutende Unterschiede für einzelne Diagnosen zwischen den theoriegeleiteten Verlaufsgruppen bestehen, generell aber die meisten Diagnosegruppen in einem ähnlichen Ausmaß über die Verlaufsgruppen verteilt sind (vgl. Abbildung 24).

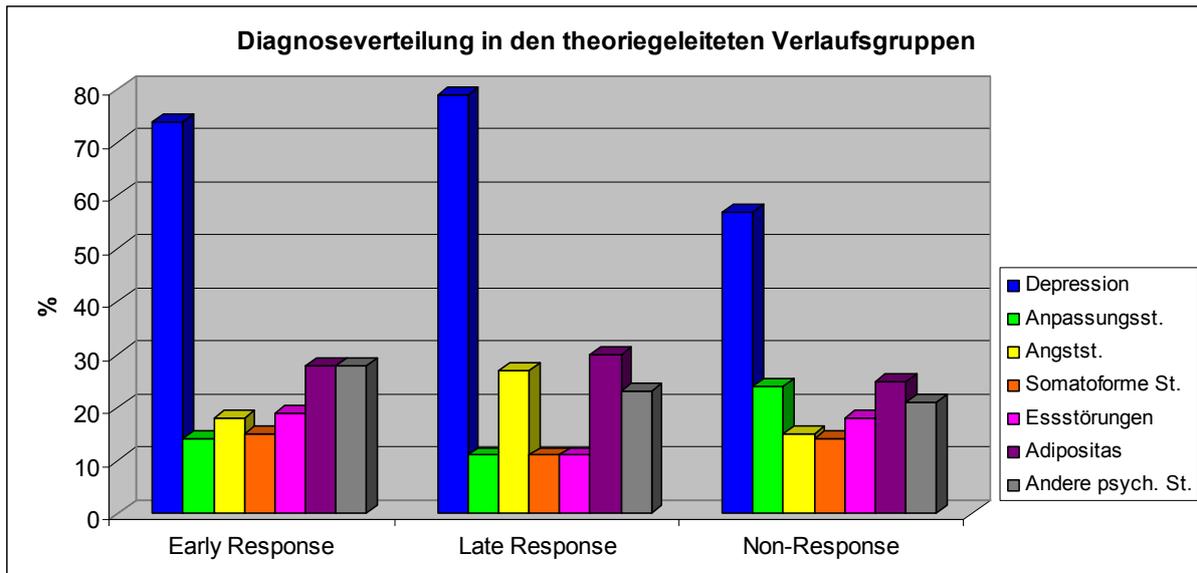


Abbildung 24: Verteilung der Diagnosegruppen in den theoriegeleiteten Verlaufgruppen

5.6.2.4 Krankheitsbezogene Merkmale

Die Symptombelastung zu Behandlungsbeginn ist bei den Patienten mit einer theoriegeleiteten Early Response_{TG} am höchsten ($M = 1.7$; $SD = 0.7$), gefolgt von der Gruppe der Late Response_{TG} ($M = 1.5$; $SD = 0.7$), und am geringsten bei den Patienten, die weder früh noch spät eine ausreichende Verbesserung erreichen (Non-Response_{TG}: $M = 0.9$; $SD = 0.8$). Der Gruppenunterschied ist hoch signifikant und von großer Effektstärke [$F(3,572) = 52.210$, $p < .001^{***}$, $\eta^2 = .215^{+++}$]. Auffällig ist hier, dass von den 112 Patienten, die zu Beginn der Behandlung einen Skalenwert unterhalb des Cut-Off-Wertes aufweisen, 70 Patienten (63%) während der Behandlung nicht respondieren. Anders gesagt liegt von den 138 theoriegeleiteten Non-Response_{TG}-Patienten die Hälfte schon zu Behandlungsbeginn in einem niedrigen Bereich der Symptombelastung. Inwiefern die Einordnung dieser Patienten in die Non-Response_{TG}-Gruppe inhaltlich und methodisch angemessen ist, muss kritisch diskutiert werden.

Entsprechend zeigen sich auch in der Skala zu *psychischen Aspekten der gesundheitsbezogenen Lebensqualität* bei Behandlungsbeginn die höchsten Werte in der Non-Response_{TG}-Gruppe (SF-8 Psych. $M = 38.3$; $SD = 13.2$), während die Gruppen Late Response_{TG} ($M = 30.2$; $SD = 11.5$) und Early Response_{TG} ($M = 28.2$; $SD = 10.1$) geringere Werte in der gesundheitsbezogenen Lebensqualität aufweisen [$F(3,561) = 29.194$, $p < .001^{***}$, $\eta^2 = .135^{++}$].

Die allgemeine Selbstwirksamkeitserwartung ist ebenfalls bei den Patienten der Non-Response_{TG}-Gruppe am höchsten ausgeprägt ($M = 25.5$; $SD = 6.9$) und unterscheidet sich signifikant in den theoriegeleiteten Verlaufgruppen (Early Response_{TG}: $M = 22.3$; $SD = 5.9$; Late Response_{TG}: $M = 21.2$; $SD = 5.6$) [$F(3,568) = 13.261$, $p < .001^{***}$, $\eta^2 = .065^{++}$].

In den sechs Skalen des Fragebogens zur Psychotherapiemotivation bestehen für die theoriegeleiteten Verlaufgruppen nur in einer Skala signifikante Gruppenunterschiede: Auf der Skala *Leidensdruck* geben die Patienten der Gruppen Early Response_{TG} ($M = 3.3$; $SD = 0.6$) und Late Response_{TG} ($M = 3.4$; $SD = 0.6$) höhere Werte an als die Patienten der Non-Response_{TG}-Gruppe ($M = 2.9$; $SD = 0.8$) [$F(3,568) = 18.082$, $p < .001^{***}$, $\eta^2 = .087^{++}$].

Zusammengefasst ist bei den theoriegeleiteten Verlaufgruppen die Non-Response_{TG}-Gruppe durch eine geringe Symptombelastung, höhere Lebensqualität, höhere Selbstwirksamkeitserwartung und einen geringeren Leidensdruck als die Gruppen Early Response_{TG} und Late Response_{TG} gekennzeichnet.

5.6.2.5 Prozessmerkmale

Die mittlere Behandlungsdauer unterscheidet sich signifikant mit mittlerer Effektstärke zwischen den theoriegeleiteten Verlaufgruppen [$F(3,572) = 24.042$, $p < .001^{***}$, $\eta^2 = .112^{++}$]. Dabei werden die Patienten der Late-Response_{TG}-Gruppe mit durchschnittlich 46 Tagen ($SD = 7$) am längsten behandelt, gefolgt von Non-Response_{TG}- ($M = 43$; $SD = 8$) und Early-Response_{TG}-Gruppe ($M = 43$; $SD = 7$). Hier muss jedoch berücksichtigt werden, dass die 90 Patienten, die aufgrund zu vieler fehlender Werte zu keiner der inhaltlich interpretierbaren theoriegeleiteten Verlaufgruppen zugeordnet werden konnten (n/a), im Durchschnitt sehr viel kürzere Behandlungszeiten mit einer größeren Streuung aufweisen ($M = 36$; $SD = 10$) und demnach ausschlaggebend für den signifikanten Gruppeneffekt sind.

Ein ähnliches Bild ist auch bei der Gesamtdosis psychotherapienaher Maßnahmen zu sehen: Der Gruppenunterschied erweist sich als signifikant [$F(3,511) = 6.874$, $p < .001^{***}$, $\eta^2 = .039^+$], wobei die n/a -Gruppe mit durchschnittlich insgesamt 27.9 Stunden die geringste Dosis erhalten hat (Gesamtdosis in Minuten: $M = 1\,674$; $SD = 707$). Die höchste Dosis mit durchschnittlich insgesamt 37.4 Stunden psychotherapienaher Behandlungsmaßnahmen haben die Patienten aus der Late-Response_{TG}-Gruppe erhalten (Gesamtdosis in Minuten: $M = 2\,242$; $SD = 929$). Die beiden Gruppen Early Response_{TG} und Non-Response_{TG} liegen bei einer durchschnittlichen Dosis von 35.2 bzw. 35.4 Stunden (Early Response_{TG}: $M = 2\,109$; $SD = 885$; Non-Response_{TG}: $M = 2\,122$; $SD = 741$). Bei einer angenommenen gleichen Verteilung der Maßnahmen über die Behandlungswochen liegt die durchschnittliche wöchentliche Dosis in der Late-Response_{TG}-Gruppe, relativiert an der durchschnittlichen Behandlungsdauer dieser Gruppe, bei 5.7 Stunden. Hier zeigt sich, dass die Gruppe n/a zwar die geringste Dosis aufweist, unter Einbezug der Behandlungsdauer mit einer Dosis von durchschnittlich 5.4 Stunden pro Woche nicht mehr so stark von der Late-Response_{TG}-Gruppe abweicht. Die Early-Response_{TG}-Gruppe weist hier die höchste wöchentliche Dosis mit Stunden 5.8 auf, gefolgt von der Non-Response_{TG}-Gruppe mit durchschnittlich 5.7 Stunden pro Woche.

Für die theoriegeleiteten Verlaufgruppen scheint die Dosis psychotherapienaher Maßnahmen weitgehend proportional zur Behandlungsdauer zu sein.

Hinsichtlich des Anteils erhaltener verhaltenstherapeutischer und psychodynamisch orientierter Behandlungsmaßnahmen bestehen keine signifikanten Unterschiede zwischen den theoriegeleiteten Verlaufgruppen [$F(3,509) = 1.340, p = .260, \eta^2 = .008$].

Die patientenseitig eingeschätzte Güte der therapeutischen Beziehung, erfasst nach der zweiten und dritten Behandlungswoche, unterscheidet sich nicht signifikant zwischen den theoriegeleiteten Verlaufgruppen.

5.6.3 Auswahl der Prädiktorvariablen

In *Hypothese 2* und *Hypothese 3* wird angenommen, dass ein Mindestanteil der Varianz in den Verlaufgruppen durch bestimmte theoretisch fundierte und empirisch abgeleitete Prädiktoren erklärt wird. Zur Untersuchung der beiden Hypothesen wurden multinomiale logistische Regressionen mit der jeweiligen Verlaufgruppenvariable als zu vorhersagendes Merkmal (abhängige Variable) berechnet. Für beide Verlaufgruppenvariablen (explorative und theoriegeleitete Verlaufgruppen) wurden getrennte Analysen berechnet.

5.6.3.1 Soziodemographische und klinische Prädiktoren des Verlaufs

Die Auswahl potentieller Prädiktoren (unabhängiger Variablen) erfolgte anhand bisheriger empirischer Befunde zur Vorhersage des Behandlungsverlaufs. Aufgrund der geringen Studienlage zur Analyse und Vorhersage von Behandlungsverläufen (vgl. Abschnitt 2.6) wurde die Vorgehensweise erweitert, indem auch Prädiktoren des Behandlungserfolgs sowie der Behandlungsdauer einbezogen wurden. Begründet ist die gewählte Herleitung möglicher Prädiktoren des Behandlungsverlaufs darin, dass der Verlauf hier in Form des Erfolgsmaßes *Symptombelastung* über die Zeit der gesamten Behandlung abgebildet wird. Es wird angenommen, dass Merkmale, die den Behandlungserfolg *nach* einer Behandlung vorhersagen (bei Entlassung oder katamnestisch), potentiell auch zur Vorhersage des Behandlungserfolgs *während* der Behandlung dienen können. Neben der literaturbasierten Auswahl der Prädiktoren wurden die Ergebnisse der univariaten Analysen (s. Abschnitt 5.6) der vorliegenden Stichprobe mit einbezogen: Merkmale, in denen signifikante Unterschiede zwischen den Verlaufgruppen bestehen, wurden in die multivariaten Analysen aufgenommen, unabhängig von ihrer literaturbasierten Fundierung. Damit soll dem explorativen Charakter der Vorhersage von Behandlungsverläufen Rechnung getragen werden. Gleichzeitig sollte bei der Auswahl darauf geachtet werden, die Anzahl der eingeschlossenen Prädiktoren möglichst gering zu halten, um eine ausreichende Teststärke zu gewährleisten. Entsprechend den Empfehlungen von Barth (1997) wurden Merkmale, zu denen die empirische Grundlage unklar bzw. unzureichend ist, und die in den univariaten

Analysen keine signifikanten Unterschiede erbrachten, nicht in die Regression mit aufgenommen.

5.6.3.2 Ausgangsbelastung als Prädiktor des Verlaufs

Die Ausgangsbelastung, operationalisiert über den Skalenwert der Symptombelastung zu Behandlungsbeginn (HEALTH-49 Modul A zu t0), wurde in die Bildung der explorativen Verlaufsgruppen direkt, also als absoluter Wert, einbezogen (vgl. Abschnitt 4.5.2). Daher wäre eine Vorhersage der Gruppenzugehörigkeit durch die Ausgangsbelastung tautologisch und wurde demnach nicht in die Regressionsanalyse zur Vorhersage der explorativen Verlaufsgruppen aufgenommen. Der Einschluss dieser Variable als Prädiktorvariable würde verhindern, potentielle Zusammenhänge anderer Merkmale mit den explorativen Verlaufsgruppen zu entdecken, was den Informationsgewinn reduzieren würde.

Für die Gruppenbildung der theoriegeleiteten Verlaufsgruppen ist die Ausgangsbelastung nur indirekt, d.h. über die Bildung von Differenzwerten zur Ausgangsbelastung, eingegangen (s. Abschnitt 4.3.4). Daher wurde der Einschluss der Ausgangsbelastung als potentieller Prädiktor zur Vorhersage der theoriegeleiteten Verlaufsgruppen als adäquat eingeschätzt.

Die in den univariaten Analysen dargestellten Zusammenhänge von Ausgangsbelastung und den Verlaufsgruppen sind in Tabelle 25 zusammengefasst dargestellt. Die Ausgangsbelastung unterscheidet sich in beiden Verlaufsgruppenvariablen jeweils hoch signifikant zwischen den einzelnen Verlaufsgruppen, der Zusammenhang zwischen Ausgangsbelastung und Verlaufsgruppe ist aber wie zu erwarten in den explorativen Verlaufsgruppen stärker ausgeprägt ($\eta^2 = .49$) als in den theoriegeleiteten Gruppen ($\eta^2 = .22$).

Dementsprechend wurde die Ausgangsbelastung in die Vorhersage der theoriegeleiteten Verlaufsgruppen einbezogen, nicht jedoch in die Vorhersage der explorativen Verlaufsgruppen (aufgrund der Tautologie).

Tabelle 25: Zusammenhang zwischen der Ausgangsbelastung (HEALTH-49 A zu t0) und Verlaufsgruppen

Verlaufsgruppen	<i>F</i> (<i>df</i>)	Signifikanz <i>p</i>	η^2
Explorative Verlaufsgruppen	139.573 (4,571)	< .001***	.494***
Theoriegeleitete Verlaufsgruppen	52.210 (3,572)	< .001***	.215***

df. Freiheitsgrade (degrees of freedom); *p.* Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (**) und große (***) Effektstärken interpretiert.

5.6.3.3 *Prozessvariablen als Prädiktoren des Verlaufs*

Die Prozessmerkmale, die sich aus den Dosisangaben anhand der KTL-Daten ableiten, weisen einen Anteil an fehlenden Werten von 11% der Verlaufsstichprobe auf. Durch den fallweisen Ausschluss bei fehlenden Werten würde die Aufnahme der Variablen in die Regressionsanalysen in einem zu hohen Datenverlust in der zur Verfügung stehenden Stichprobe resultieren. Die Power der Analysen wäre unzureichend, sodass diese Variablen nicht in die Analysen eingeschlossen wurden. Die Behandlungsdauer wurde als Prädiktor der Verlaufgruppen untersucht. Da der Zusammenhang zwischen der Behandlungsdauer und der Dosis psychotherapienaher Behandlungsmaßnahmen in einer mittleren Größenordnung liegt (Pearsons $r = .38$; s. Anhang: Tabelle A.4), wird der Verzicht auf die Analyse der Dosis als Prädiktor als akzeptabel eingeschätzt.

Da die Erhebung der Behandlungsdauer, anders als die oben genannten Patientenmerkmale, zeitlich erst *nach* der Erhebung des Symptomverlaufs möglich ist, und eine Anpassung der Dauer an den Symptomverlauf nicht ausgeschlossen ist, kann die Untersuchung dieser Prozessvariable hier lediglich im Sinne eines Zusammenhangs, und nicht einer Vorhersage, mit den Verlaufgruppen verstanden werden. Die Frage, inwiefern der Symptomverlauf aus Prozessmerkmalen vorhergesagt bzw. inwiefern dabei auf Kausalzusammenhänge geschlossen werden kann, sei an dieser Stelle schon erwähnt und muss kritisch diskutiert werden.

5.6.3.4 *Übersicht über die untersuchten Prädiktoren des Verlaufs*

Nach der Berücksichtigung literaturbasierter Merkmale und der Ergebnisse aus der vorliegenden Stichprobe lassen sich die in Tabelle 26 beschriebenen Merkmale identifizieren, die in die Analyse aufgenommen werden. Aufgrund der geringen Zellbesetzungen der Merkmalsausprägungen in den soziodemographischen Variablen Partnersituation, Schulabschluss, Berufsabschluss und Erwerbstätigkeit (s. Anhang: Tabelle C.1 und Tabelle D.1), wurden diese zu übergeordneten Kategorien zusammengefasst. Die jeweiligen Antwortkategorien bzw. Operationalisierung sind ebenfalls in Tabelle 26 aufgeführt. Die univariaten Analysen zur Untersuchung von Gruppenunterschieden in den explorativen bzw. theoriegeleiteten Verlaufgruppen sind im Anhang zu finden (explorative Verlaufgruppen: Tabelle C.1 bis Tabelle C.7; theoriegeleitete Verlaufgruppen: Tabelle D.1 bis Tabelle D.7).

Tabelle 26: Auswahl der Prädiktoren zur Vorhersage der Verlaufgruppen

Merkmal (potentieller Prädiktor)	Lit. bas.	EG sign	TG sign	log. Reg	Operationalisierung des Merkmals für die logistische Regression
Patientenmerkmale					
Geschlecht	x	--	--	x	
Alter	x	--	--	x	in Jahren
Partnersituation	x	--	--	x	in Partnerschaft (verheiratet oder ledig) / ohne Partner (ledig, geschieden, verwitwet oder getrennt lebend)
Schulabschluss	x	x	--	x	hoch (Realschule/Abitur, Hochschulabschluss) / niedrig (Hauptschule, kein Schulabschluss)
Berufsabschluss ¹	?	x	--	--	hoch (Meister, FH, Uni) / niedrig (kein, Lehre, andere)
Erwerbstätigkeit	x	x	(x) ²	x	erwerbstätig / nicht erwerbstätig
Haushaltseinkommen	--	--	--	--	in fünf Einkommenskategorien
Arbeitsunfähigkeitszeiten	x	x	--	x	in Wochen
Stationäre Vorbehandlung	?	x	--	x	ja / nein
Ambulante Vorbehandlung	?	--	--	x	ja / nein
Unipolare Depressionen ³	x	--	x	x	vorhanden / nicht vorhanden
Anpassungsstörungen ³		--	x	x	vorhanden / nicht vorhanden
Angststörungen ³		x	--	x	vorhanden / nicht vorhanden
Somatoforme Störungen ³		x	x	x	vorhanden / nicht vorhanden
PTBS ³		x	--	x	vorhanden / nicht vorhanden
Essstörungen ³		x	--	x	vorhanden / nicht vorhanden
Persönlichkeitsst. ³		x	--	x	vorhanden / nicht vorhanden
Adipositas ³		--	--	x	vorhanden / nicht vorhanden
Sonstige F-Diagnosen ³		x	--	--	vorhanden / nicht vorhanden
Psychische Komorbidität	x	x	x	x	mehr als eine F-Diagnose / eine F- Diagnose
Psychotherapie- motivation Psychischer Leidensdruck	x	x	x	x	FPTM – 23: Skala <i>Psychischer Leidensdruck</i>
Psychotherapie- motivation Hoffnung	x	x	--	x	FPTM – 23: Skala <i>Hoffnung</i>
Psychotherapie- motivation Verneinung psych. Hilfsbedürftigkeit	x	x	--	x	FPTM – 23: Skala <i>Verneinung psychischer Hilfsbedürftigkeit</i>
Psychotherapie- motivation Wissen	--	--	--	--	FPTM – 23: Skala <i>Wissen</i>
Psychotherapie- motivation Initiative	--	--	--	--	FPTM – 23: Skala <i>Initiative</i>
Psychotherapie- motivation Symptombezogene Zuwendung	x	x	--	x	FPTM – 23: Skala <i>Symptombezogene Zuwendung</i>

Merkmal (potentieller Prädiktor)	Lit. bas.	EG sign	TG sign	log. Reg	Operationalisierung des Merkmals für die logistische Regression
Allgemeine Selbstwirksamkeitserwartung	?	x	x	x	SWE Summenskala
Therapeutische Beziehung in der frühen Phase der Behandlung	x	x	--	x	HAQ: Skala <i>Beziehungszufriedenheit</i> zu Woche 2-3
Symptombelastung zu t0	x	x	x	(x) ⁴	HEALTH-49 A zu t0
Prozessmerkmale					
Behandlungsdauer	?	x	x	x	in Tagen
Dosis Psychotherapie	?	x	x	(x) ⁵	in Minuten
Dosis therapieschulenspezifischer Psychotherapie	?	--	--	(x) ⁵	Prozentualer Anteil verhaltenstherapeutischer bzw. psychodynamischer Behandlungsmaßnahmen
Behandelnde Klinik	--	--	x	x	

Li. bas.: Literaturbasiert; EG: Explorative Verlaufsgruppen; TG: Theoriegeleitete Verlaufsgruppen; log. Reg.: in die logistische Regression einbezogen

Anmerkung: ¹Der Einschluss der Variable *Berufsabschluss* in die Regression würde aufgrund der hohen Anzahl fehlender Werte zu einem zu hohen Datenverlust und damit einer Senkung der Teststärke führen. Diese Variable wird auch aus statistischen Gründen ausgeschlossen, was akzeptabel erscheint, da der Berufs- bzw. Bildungsstand durch die beiden Variablen *Schulabschluss* und *Erwerbstätigkeit* ausreichend abgebildet wird.

² Bei der Reduktion des Merkmals auf zwei Kategorien (erwerbstätig vs. nicht erwerbstätig) besteht kein signifikanter Unterschied mehr zwischen den theoriegeleiteten Verlaufsgruppen.

³ Die Diagnosegruppen wurden für jeden Patienten aus allen vergebenen Diagnosen gebildet.

⁴ Die Symptombelastung zu Behandlungsbeginn wird aus methodischen Gründen für die Vorhersage der theoriegeleiteten Verlaufsgruppen einbezogen, nicht aber für die Vorhersage der explorativen Verlaufsgruppen

⁵ Die Prozessmerkmale *Dosis psychotherapienah* und *Dosis schulenspezifisch* werden aufgrund einer zu hohen Anzahl fehlender Werte, die mit einer Senkung der Teststärke des Gesamtmodells einhergehen würde, nicht in die logistischen Regressionsanalysen der 2. Fragestellung aufgenommen. Für die linearen Regressionen der 3. Fragestellung konnten beide Variablen eingeschlossen werden.

Insgesamt wurden demnach 25 potentielle Prädiktoren der Verlaufsgruppen untersucht: die fünf soziodemographischen Variablen Alter, Geschlecht, Partnersituation, Schulabschluss und Erwerbstätigkeit, die drei sozialmedizinischen Merkmale Arbeitsunfähigkeitszeiten, ambulante Vorbehandlung und stationäre Vorbehandlung, die acht Diagnosegruppen unipolare Depressionen, Anpassungsstörungen, Angststörungen, Somatoforme Störungen, Posttraumatische Belastungsstörungen, Essstörungen, Persönlichkeitsstörungen und Adipositas, sowie die psychische Komorbidität, die vier Skalen *Leidensdruck*, *Hoffnung*, *Verneinung psychischer Hilfsbedürftigkeit* und *symptombezogene Zuwendung* der Psychotherapiemotivation, die Skala der *allgemeinen Selbstwirksamkeitserwartung*, die Skala zur Güte der therapeutischen Beziehung, die Behandlungsdauer sowie die behandelnde Klinik.

5.6.4 Vorhersage der Verlaufgruppen

5.6.4.1 Prüfung der Voraussetzungen der Vorhersage

Eine Voraussetzung linearer und logistischer Regressionen besteht darin, dass keine lineare Abhängigkeit zwischen den unabhängigen Variablen (Regressoren) bestehen sollte. Ist diese Voraussetzung verletzt, so spricht man von *Multikollinearität*. Im Fall von Multikollinearität überschneiden sich die Streuungen der unabhängigen Variablen, was zum einen zu redundanter Information, zum anderen zu Problemen bei der Interpretierbarkeit der Ergebnisse führt. Korrelieren zwei Merkmale hoch miteinander, kann bei einem signifikanten Einfluss auf die Varianz der abhängigen Variablen nicht zurückgeführt werden, welches der beiden Merkmale mit welchem Anteil zur Varianzaufklärung beiträgt. Zur Prüfung auf Multikollinearität wird in einem ersten Schritt die Berechnung einer Korrelationsmatrix empfohlen, bei der die Höhe des Zusammenhangs zwischen den Merkmalen aufgedeckt werden kann. Die Korrelationsmatrizen, in denen alle Zusammenhangsmaße der Regressoren mit Angabe der Signifikanz aufgeführt sind, finden sich im Anhang in Tabelle A.1 bis Tabelle A.6. Hier zeigt sich, dass viele Variablen zwar signifikant miteinander zusammenhängen, die Stärke des Zusammenhangsmaß (Pearsons r , Phi, Cramers V , Eta) aber meist im kleinen Bereich liegt. Keines der Zusammenhangsmaße erreicht die Höhe einer großen Effektstärke, was ein erster Hinweis auf das Nichtvorliegen von Multikollinearität ist.

Da die Korrelationskoeffizienten nur die paarweisen Abhängigkeiten abbilden, sollten zur Aufdeckung von Multikollinearität in einem zweiten Schritt Regressionen, in denen jede einzelne unabhängigen Variablen X_j aus den anderen unabhängigen Variablen vorhergesagt wird, durchgeführt werden (Backhaus et al., 2003). Das Bestimmungsmaß der Regression der unabhängigen Variablen auf die übrigen unabhängigen Variablen stellt dabei das R_j^2 dar, als Kennwert der Multikollinearität wird die sog. *Toleranz* berechnet ($T_j = 1 - R_j^2$). Der Kehrwert der Toleranz ist die sog. *Varianzinflationsfaktor* (*Variance Inflation Factor*, $VIF_j = 1/T_j$). Entsprechen liegt bei niedrigen Toleranzwerten ($T_j < 0.1$) und hohen Werten des Varianzinflationsfaktors ($VIF_j > 10$) ein Hinweis auf Multikollinearität vor (Brosius, 2008; Menard, 2010). Die Toleranzwerte liegen in den vorliegenden Daten zwischen 0.33 und 0.92 und der *VIF* zwischen 1.1 und 3.0 und deuten damit nicht auf verdeckte Multikollinearität hin²⁸.

²⁸ Die Berechnung der Multikollinearitäts-Statistik ist mit der Software SPSS 15.0 nur bei der linearen Regression möglich. Zur Berechnung der Multikollinearitäts-Kennwerte für die logistische Regression wurde zusätzlich eine lineare Regression mit den gleichen Variablen durchgeführt, da die Modellstatistik der abhängigen Variable für diesen Zweck irrelevant ist und die benötigten Koeffizienten der unabhängigen Variablen berechnet werden können (Menard, 2010).

Eine weitere Voraussetzung der Regression besteht darin, dass die Residuen, also die Fehler der Schätzungen, keinem systematischen Muster entsprechen und zufällig auftreten. Mithilfe des Durbin-Watson-Koeffizienten kann die sog. *Autokorrelation der Residuen* getestet werden. Ein Koeffizient, der deutlich kleiner als 2 ist, zeigt eine positive Autokorrelation der Residuen an, ein Wert deutlich größer als 2 eine negative Autokorrelation. Werte zwischen 1.5 und 2.5 werden als akzeptabel eingestuft (Brosius, 2008). Der Durbin-Watson-Koeffizient des vorliegenden Modells liegt bei 2.17, sodass auch die Voraussetzung der unsystematischen Residuen weitgehend erfüllt ist.

5.6.4.2 Ergebnisse zur Vorhersage der explorativen Verlaufsguppen (Fragestellung 2 A)

Die Ergebnisse der logistischen Regression mit der Gruppenvariable der explorativen Verlaufsguppen als abhängige Variable und den in Tabelle 26 (vgl. Abschnitt 5.6.3.4; S. 143) aufgeführten Merkmalen als unabhängige Variablen sind in Tabelle 27 bis Tabelle 29 dargestellt.

Tabelle 27: Modellzusammenfassung der logistischen Regression zur Vorhersage der Gruppenzugehörigkeit zu den *explorativen Verlaufsguppen*

<i>n</i> Gesamt	Chi ² (<i>df</i>)	Signifikanz <i>p</i>	Pseudo- <i>R</i> ² (Nagelkerke)
514	290.572 (104)	< .001***	.503

df: Freiheitsgrade (degrees of freedom); *p*: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Insgesamt konnten $n = 514$ Patienten in die logistische Regressionsanalyse aufgenommen werden. Die Güte des Gesamtmodells ist hoch signifikant und mit einem Pseudo- R^2 von .50 als sehr gut zu bewerten. Das bedeutet, dass 50% der Varianz in der abhängigen Variable, hier die explorativen Verlaufsguppen, durch das untersuchte Set an unabhängigen Variablen erklärt werden. Tabelle 28 enthält die statistischen Kennwerte der über das Gesamtmodell signifikanten Prädiktoren, Tabelle 29 die Kennwerte der jeweiligen Verlaufsguppenvergleiche. Dabei werden aufgrund des explorativen Charakters der Fragestellung auch tendenziell signifikante Prädiktoren ($\alpha < .10$) dargestellt (s. Abschnitt 4.5.4).

Tabelle 28: Signifikante Prädiktoren der logistische Regression; Zielkriterium *explorative Verlaufgruppen*

Prädiktor	Chi ² (df)	Signifikanz p
Behandlungsdauer (Tage)	9.809 (4)	.044*
Arbeitsunfähigkeitszeiten (Tage)	13.611 (4)	.009**
FPTM Leidensdruck	27.341 (4)	< .001***
Selbstwirksamkeitserwartung	9.669 (4)	.046*
Schulabschluss	18.738 (4)	.001**
Erwerbstätigkeit	9.156 (4)	.057°
Essstörungen	8.119 (4)	.087°
psychische Komorbidität	9.101 (4)	.059°

df. Freiheitsgrade (degrees of freedom); p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt.

Über das Gesamtmodell erweisen sich von 25 untersuchten Prädiktoren die 5 Merkmale Behandlungsdauer, Arbeitsunfähigkeitszeiten, Leidensdruck, Selbstwirksamkeitserwartung und Schulabschluss als signifikant und des Weiteren die 3 Merkmale Erwerbstätigkeit, Diagnose Essstörung und psychische Komorbidität als tendenziell signifikant.

Die Parameterschätzer aus den paarweisen Vergleichen sind in Tabelle 29 dargestellt. Als Referenzgruppe für die Vergleiche wurde die Gruppe *Linear Response_{EG}* gewählt, zum einen da diese Gruppe aufgrund der Gruppengröße ($n = 367$) inhaltlich als Standardgruppe gelten kann und zum anderen infolgedessen die statistische Power der Einzelvergleiche erhöht wird.

Tabelle 29: Odds Ratios, 95% Konfidenzintervall und Signifikanz der logistischen Regression; Zielkriterium *explorative Verlaufgruppen* (Referenzgruppe: Linear Response_{EG}, n = 367)

Gruppe	Prädiktor	OR [95% CI]	p
Early Response_{EG}			
(n = 46)	FPTM Leidensdruck	2.82 [1.27; 6.25]	.011*
	FPTM Verneinung ¹	1.64 [0.99; 2.72]	.054°
	Selbstwirksamkeitserwartung	0.93 [0.87; 1.00]	.042*
	Somatoforme Störung	2.38 [0.87; 6.49]	.090°
	Psychische Komorbidität	2.80 [0.96; 8.13]	.059°
Delayed Response_{EG}			
(n = 26)	Alter	1.05 [1.00; 1.10]	.078°
	Arbeitsunfähigkeitszeiten	1.06 [1.00; 1.12]	.049*
	FPTM Leidensdruck	12.58 [2.57; 61.62]	.002**
	Schulabschluss niedrig	2.62 [0.92; 7.43]	.070°
	Persönlichkeitsstörung	7.35 [1.51; 35.71]	.014*
Non-Response_{EG}			
(n = 54)	Arbeitsunfähigkeitszeiten	1.07 [1.03; 1.11]	.001***
	FPTM Leidensdruck	4.20 [1.86; 9.46]	.001***
	FPTM Verneinung ¹	1.60 [0.95; 2.69]	.075°
	FPTM Sympt. Zuwendung ¹	0.65 [0.39; 1.08]	.094°
	Bestehende Partnerschaft	2.23 [1.06; 4.63]	.034*
	Schulabschluss niedrig	4.57 [2.12; 9.85]	< .001***
	Nicht erwerbstätig	2.34 [1.09; 5.02]	.029*
	Ambulante Vorbehandlung	0.37 [0.17; 0.79]	.011*
Psychische Komorbidität	4.02 [1.28; 12.66]	.018*	
Atypical Course			
(n = 21)	Behandlungsdauer	0.94 [0.89; 1.00]	.035*
	FPTM Hoffnung	0.38 [0.16; 0.92]	.031*
	Selbstwirksamkeitserwartung	0.92 [0.83; 1.02]	.097°
	Schulabschluss niedrig	3.01 [1.02; 8.94]	.047*
	Essstörung	5.99 [1.41; 25.64]	.015*

OR: Odds Ratio; CI: Konfidenzintervall; p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Für jede der vier Verlaufgruppen wurden alle 25 eingeschlossenen Prädiktoren getestet: Alter, Geschlecht, Partnersituation, Schulabschluss, Erwerbstätigkeit, Arbeitsunfähigkeitszeiten, ambulante und stationäre Vorbehandlung, Unipolare Depressionen, Anpassungsstörungen, Angststörungen, Somatoforme Störungen, Posttraumatische Belastungsstörungen, Essstörungen, Persönlichkeitsstörungen, Adipositas, psychische Komorbidität, Leidensdruck, Hoffnung, Verneinung psychischer Hilfsbedürftigkeit, symptombezogene Zuwendung, allgemeine Selbstwirksamkeitserwartung, Güte der therapeutischen Beziehung, Behandlungsdauer und Klinik. Zwecks Übersichtlichkeit werden hier nur die signifikanten und tendenziell signifikanten Prädiktoren jeder Gruppe dargestellt. Die dichotomen Prädiktoren sind jeweils in Richtung des Vorhandenseins des Merkmals analysiert worden (Bsp. Essstörung im Sinne des Vorhandenseins der Diagnose).

¹ Die FPTM-Skalen *Verneinung psychischer Hilfsbedürftigkeit* und *Symptombezogene Zuwendung* sind negativ gepolt. Ein hoher Skalenwert zeigt entsprechend eine geringe Psychotherapiemotivation an.

Das Odds Ratio (OR) gibt für den untersuchten Prädiktor an, wie sich das Chancenverhältnis für die Gruppenzugehörigkeit ändert, wenn der empirische Wert des Prädiktor um eine

Einheit erhöht wird. Bei der Interpretation der *OR*-Werte muss die Skalierung des jeweiligen Prädiktors berücksichtigt werden.

Es zeigt sich, dass Patienten eher ein frühes Ansprechen aufweisen (*höhere Chance*²⁹ *der Zugehörigkeit zur Early-Response_{EG}-Gruppe*), wenn sie höhere Werte auf den Psychotherapiemotivation-Skalen *Leidensdruck* [*OR* = 2.82, *CI*: 1.27 - 6.25, *p* = .011] und *Verneinung psychischer Hilfsbedürftigkeit*³⁰ [*OR* = 1.64, *CI*: 0.99 - 2.72, *p* = .054] aufweisen, wenn sie die (Ko-)Diagnose einer somatoformen Störung [*OR* = 2.38, *CI*: 0.87 - 6.49, *p* = .090] und generell mehr als eine F-Diagnose [*OR* = 2.80, *CI*: 0.96 - 8.13, *p* = .059] aufweisen. Mit höherer Selbstwirksamkeitserwartung zu Behandlungsbeginn sinkt die Chance, in die Gruppe der Early Response_{EG} zu gelangen [*OR* = 0.93, *CI*: 0.87 - 1.00; *p* = .042].

Allerdings liegen die Werte des Konfidenzintervalls nur für den Prädiktor *Leidensdruck* außerhalb des Wertes des Null-Effektes (für *OR*: 1), sodass nur bei diesem Prädiktor von einem auf dem 5%-Niveau abgesicherten Odds Ratio gesprochen werden kann [*OR* = 2.82, *CI*: 1.27 - 6.25, *p* = .011]. Patienten mit einem um eine Einheit erhöhten *Leidensdruck* haben demnach mit einer 95%igen Vertrauenswahrscheinlichkeit eine 2.8fach höhere Chance eine Early Response_{EG} aufzuweisen. Die itemstandardisierte Skala *Leidensdruck* des FPTM-23 umfasst den Wertebereich von 1 bis 4, sodass ein Anstieg um den Wert 1 schon ein Drittel der Skala ausmacht.

Im Vergleich dazu spiegelt eine Erhöhung um den Wert 1 auf der SWE-Skala eine geringere Ausprägung wider, da der Wertebereich dieser Skala zwischen 10 und 40 liegt (Summenscore). Das *OR* der Selbstwirksamkeitserwartung liegt bei 0.93, was bedeutet, dass sich bei einem Anstieg der Selbstwirksamkeitserwartung um den Wert 1 das Chancenverhältnis um 0.93 erhöht. Das ist eine geringere Abweichung der Chancengleichheit eines *OR* von 1, bei Berücksichtigung des breiteren Skalenranges relativiert sich dieser Unterschied. Eine Erhöhung des Wertes 1 auf der Skala *Leidensdruck* würde einer Erhöhung um 10 Einheiten auf der Selbstwirksamkeitsskala entsprechen. Das *OR* für das Chancenverhältnis bei einem um das 10fache erhöhten SWE-Skalenwert würde entsprechend bei $0.93^{10} = 0.48$ liegen.

Eine *erhöhte Chance einer Delayed Response_{EG}* zeigt sich signifikant, d.h. mit einem Konfidenzintervall außerhalb des Wertes 1, für Patienten mit höheren Werten in der Skala *Leidensdruck* [*OR* = 12.58, *CI*: 2.57 - 61.62, *p* = .002] und der (Ko-)Diagnose einer Persönlichkeitsstörung [*OR* = 7.35, *CI*: 1.51 - 35.71, *p* = .014]. Tendenziell signifikant werden

²⁹ Es ist zu beachten, dass der Begriff der Chance unabhängig davon verwendet wird, ob das Zielereignis positiv oder negativ ist. Auch bei einem negativen Zielereignis (bspw. einer Non-Response) spricht man von der Chance der Gruppenzugehörigkeit.

³⁰ Höhere Werte auf der Skala *Verneinung psychischer Hilfsbedürftigkeit* zeigen eine geringe Psychotherapiemotivation an (vgl. Abschnitt 4.3.3.2).

die Effekte für eine erhöhte Chance bei längeren Arbeitsunfähigkeitszeiten [$OR = 1.06$, $CI: 1.00 - 1.12$, $p = .049$], einem niedrigen Schulabschluss [$OR = 2.62$, $CI: 0.92 - 7.43$, $p = .070$] und höherem Alter [$OR = 1.05$, $CI: 1.00 - 1.10$, $p = .078$]. Es kann nicht mit 95%iger Vertrauenswahrscheinlichkeit gesagt werden, dass der angezeigten Effektstärken der drei letztgenannten Prädiktoren tatsächlich eintreten.

Das bedeutet, dass Patienten mit einem um eine Einheit erhöhten Wert auf der Skala *Leidensdruck* eine 12.6fach erhöhte Chance einer Delayed Response_{EG} aufweisen. Patienten mit einer (singulären oder komorbiden) Persönlichkeitsstörung haben eine 7.4fach erhöhte Chance einer Delayed Response_{EG}, verglichen mit Patienten, bei denen keine Persönlichkeitsstörung diagnostiziert wurde. Diese beiden Prädiktoren weisen die höchsten OR Werte auf, das 95%ige Konfidenzintervall liegt jeweils außerhalb des Null-Effektes. Eine geringere Zuverlässigkeit weisen die Prädiktoren niedriger Schulabschluss, der die Chance einer Delayed Response_{EG} um das 2.6fache erhöht; Alter, das mit jedem Lebensjahr die Chance einer Delayed Response_{EG} um das 1.05fache erhöht; und Arbeitsunfähigkeitszeiten, die mit jedem zusätzlichen Tag der Krankschreibung die Chance einer Delayed Response_{EG} um das 1.06fache erhöhen, auf.

Eine *höhere Chance für die Zugehörigkeit zur Non-Response_{EG}-Gruppe* geht einher mit niedrigem Schulabschluss [$OR = 4.57$, $CI: 2.12 - 9.85$, $p < .001$], längeren Arbeitsunfähigkeitszeiten [$OR = 1.07$, $CI: 1.03 - 1.11$, $p = .001$], psychischer Komorbidität [$OR = 4.02$, $CI: 1.28 - 12.66$, $p = .018$], einer Nichterwerbstätigkeit [$OR = 2.34$, $CI: 1.09 - 5.02$, $p = .029$], einer bestehenden Partnerschaft [$OR = 2.23$, $CI: 1.06 - 4.63$, $p = .034$] und einem höheren Werten auf der Skala *Leidensdruck* [$OR = 4.20$, $CI: 1.86 - 9.46$, $p = .001$]. Patienten mit ambulanter psychotherapeutischer Vorbehandlung weisen eine geringere Chance einer Non-Response_{EG} im Behandlungsverlauf auf [$OR = 0.37$, $CI: 0.17 - 0.79$, $p = .011$].

Die beiden Skalen *Verneinung psychischer Hilfsbedürftigkeit* [$OR = 1.60$, $CI: 0.95 - 2.69$, $p = .075$] und *Symptombezogene Zuwendung* [$OR = 0.65$, $CI: 0.39 - 1.08$, $p = .094$] zeigen einen tendenziell signifikanten Zusammenhang. Höhere Werte in der Verneinung, die auf eine niedrige Psychotherapiemotivation hindeuten, und geringere Werte in der symptombezogenen Zuwendung, die eine hohen Psychotherapiemotivation anzeigen, verändern das Chancenverhältnis zugunsten einer Non-Response_{EG}. Allerdings schließen die Konfidenzintervalle der OR beider Prädiktoren den Wert 1 des Null-Effektes mit ein, sodass diese Ergebnisse vorsichtig interpretiert werden sollten. Die vergleichsweise niedrige Ausprägung des OR des Prädiktors Arbeitsunfähigkeitszeiten ist durch den breiteren Skalenrange erklärbar: Die Arbeitsunfähigkeitszeiten wurden in Tagen erfasst, sodass eine Krankschreibung, die um einen Tag erhöht ist, mit einem Chancenverhältnis von 1.07: 1 zugunsten der Non-Response_{EG} einhergeht. Zur Veranschaulichung, was dieses für den

Zeitraum einer Woche ausmacht, sei folgende Berechnung angegeben: War ein Patient vor Behandlungsbeginn eine Woche länger krankgeschrieben, so erhöht sich die Chance einer Non-Response_{EG} schon auf das $1.07^7 = 1.61$ fache. Anhand der Potenzierung der OR lassen sich Profile für bestimmte Risikogruppen bilden: So hätte beispielsweise ein nicht erwerbstätiger Patient mit niedrigem Schulabschluss und mehr als einer F-Diagnose demnach eine 43fach erhöhte Chance einer Non-Response_{EG} (OR: $4.57 \cdot 4.02 \cdot 2.34$). Die Parameterschätzer der Gruppe Atypical Course werden nicht interpretiert.

Hinsichtlich der zu Fragestellung 2 A postulierten Hypothese zeigen die Ergebnisse, dass durch die theoretisch und empirisch fundierten Prädiktoren ein signifikanter Varianzanteil in den explorativen Verlaufgruppen erklärt werden kann. Die *Hypothese 2* wird angenommen. Die Nullhypothese kann abgelehnt werden, da die Modellgüte sich signifikant von Null unterscheidet.

Insgesamt werden 50% der Varianz in den explorativen Verlaufgruppen durch die untersuchten Patientenmerkmale erklärt, was der Höhe einer großen Effektstärke (Pseudo- $R^2 \geq .50$; vgl. Tabelle 3, Abschnitt 4.5.4) entspricht.

Wird zur Vorhersage der explorativen Verlaufgruppen zusätzlich die Ausgangsbelastung (HEALTH-49 A zu t0) in das Modell mit einbezogen, erhöht sich die erklärte Varianz der explorativen Verlaufgruppen auf 71% [$\text{Chi}^2 = 482.755$; $df = 108$; $p < .001$; Nagelkerkes Pseudo- $R^2 = .710$].

5.6.4.3 Vorhersage der theoriegeleiteten Verlaufgruppen (Fragestellung 2 B)

Die Ergebnisse der logistischen Regression mit der Gruppenzugehörigkeit der theoriegeleiteten Gruppen als abhängige Variable und den in Tabelle 26 (Abschnitt 5.6.3.4; S. 143) aufgeführten Merkmalen als unabhängige Variablen sind in Tabelle 30 bis Tabelle 32 dargestellt.

Tabelle 30: Modellzusammenfassung der logistischen Regression zur Vorhersage der Gruppenzugehörigkeit zu den *theoriegeleiteten Verlaufgruppen*

n Gesamt	Chi ² (df)	Signifikanz p	Pseudo-R ² (Nagelkerke)
514	319.315 (81)	< .001***	.508

df: Freiheitsgrade (degrees of freedom); p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Hier zeigt sich ebenfalls eine sehr gute Modellgüte mit einem Pseudo- R^2 von .52. Die aufgeklärte Varianz der theoriegeleiteten Gruppen durch die in die Regressionsanalyse eingeschlossenen Prädiktoren beträgt 51%. Tabelle 31 und Tabelle 32 enthalten die

statistischen Kennwerte der signifikanten Prädiktoren des Gesamtmodells sowie die Parameterschätzer der Einzelvergleiche zwischen den theoriegeleiteten Gruppen.

Tabelle 31: Signifikante Prädiktoren der logistische Regression; Zielkriterium *theoriegeleitete Verlaufgruppen*

Prädiktor	Chi ² (df)	Signifikanz p
Behandlungsdauer	63.751 (3)	< .001***
Arbeitsunfähigkeitszeiten	12.057 (3)	.007**
Therapeutische Beziehung	20.984 (3)	< .001***
Selbstwirksamkeitserwartung	7.182 (3)	.066°
Ausgangsbelastung (HEALTH-49 A t0)	103.320 (3)	< .001***
Schulabschluss	7.970 (3)	.047*
Ambulante Vorbehandlung	8.501 (3)	.037*
Angststörung	6.737 (3)	.081°
Persönlichkeitsstörung	8.852 (3)	.031*
Klinik	27.462 (6)	< .001***

df. Freiheitsgrade (degrees of freedom); p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Im Gesamtmodell werden 8 der 26 untersuchten Prädiktoren signifikant, 2 weitere weisen eine Tendenz auf. Die Klinik erweist sich über das Gesamtmodell als hoch signifikant und wird bei den Ergebnissen der Parameterschätzer genauer betrachtet. Als Referenzgruppe wurde hier die Gruppe *Early Response_{TG}* gewählt, da diese Gruppe aufgrund der Gruppengröße ($n = 259$) inhaltlich am ehesten als Standardgruppe gelten kann und infolgedessen die statistische Power der Einzelvergleiche erhöht wird.

Tabelle 32: Odds Ratios, 95% Konfidenzintervall und Signifikanz der logistischen Regression; Zielkriterium *theoriegeleitete Verlaufsgruppen* (Referenzgruppe: Early Response_{TG}, n = 259)

Gruppe	Prädiktor	OR [95% CI]	p
Late Response_{TG}			
(n = 52)	Arbeitsunfähigkeitszeiten	1.06 [1.02; 1.11]	.006**
	Selbstwirksamkeitserwartung	0.94 [0.88; 1.01]	.063°
	Ausgangsbelastung	0.43 [0.23; 0.82]	.010*
	Angststörung	3.02 [1.14; 8.00]	.026*
	Persönlichkeitsstörung	3.44 [0.85; 13.89]	.084°
	Adipositas	2.25 [0.91; 5.59]	.079°
	Psychische Komorbidität	0.40 [0.14; 1.14]	.087°
Non-Response_{TG}			
(n = 120)	Behandlungsdauer	1.04 [1.01; 1.08]	.022*
	Arbeitsunfähigkeitszeiten	1.06 [1.02; 1.09]	.001***
	Therapeutische Beziehung	0.88 [0.83; 0.94]	< .001***
	Ausgangsbelastung	0.06 [0.03; 0.11]	< .001***
	Bestehende Partnerschaft	1.68 [0.94; 3.02]	.080°
	Ambulante Vorbehandlung	2.08 [1.16; 3.75]	.014*
	Somatoforme Störung	2.49 [0.94; 6.58]	.066°
	Persönlichkeitsstörung	5.59 [1.65; 18.87]	.006**
n/a			
(n = 83)	Behandlungsdauer	0.89 [0.86; 0.93]	< .001***
	FPTM Sympt. Zuwendung ¹	0.66 [0.44; 0.98]	.040*
	Therapeutische Beziehung	0.91 [0.85; 0.97]	.002**
	Ausgangsbelastung	0.38 [0.21; 0.71]	.002**
	Niedriger Schulabschluss	2.64 [1.29; 5.39]	.008**
	Angststörung	2.25 [0.88; 5.78]	.091°
	Klinik 1	0.25 [0.09; 0.70]	.008**
	Klinik 2	2.50 [1.09; 5.71]	.030*

OR: Odds Ratio; CI: Konfidenzintervall; p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Für jede der drei Verlaufsgruppen wurden alle 26 eingeschlossenen Prädiktoren getestet: Ausgangsbelastung, Alter, Geschlecht, Partnersituation, Schulabschluss, Erwerbstätigkeit, Arbeitsunfähigkeitszeiten, ambulante und stationäre Vorbehandlung, Unipolare Depressionen, Anpassungsstörungen, Angststörungen, Somatoforme Störungen, Posttraumatische Belastungsstörungen, Essstörungen, Persönlichkeitsstörungen, Adipositas, psychische Komorbidität, Leidensdruck, Hoffnung, Verneinung psychischer Hilfsbedürftigkeit, symptombezogene Zuwendung, allgemeine Selbstwirksamkeitserwartung, Güte der therapeutischen Beziehung, Behandlungsdauer und Klinik. Zwecks Übersichtlichkeit werden hier nur die signifikanten und tendenziell signifikanten Prädiktoren jeder Gruppe dargestellt. Die dichotomen Prädiktoren sind jeweils in Richtung des Vorhandenseins des Merkmals analysiert worden (Bsp. Essstörung im Sinne des Vorhandenseins der Diagnose).

¹ Die FPTM-Skalen *Verneinung psychischer Hilfsbedürftigkeit* und *Symptombezogene Zuwendung* sind negativ gepolt. Ein hoher Skalenwert zeigt entsprechend eine geringe Psychotherapiemotivation an.

Bei der *Vorhersage der Gruppenzugehörigkeit zu der Late-Response_{TG}-Gruppe* wurden die Prädiktoren Arbeitsunfähigkeitszeiten, Ausgangsbelastung und die Diagnose Angststörung signifikant. Patienten mit längeren Arbeitsunfähigkeitszeiten haben eine 1.06fach erhöhte Chance einer Late Response_{TG} [OR = 1.06, CI: 1.02 – 1.11, p = .006], Patienten mit der (Ko-

)Diagnose einer Angststörung haben eine 3.02fach erhöhte Chance einer Late Response_{TG}. Eine höhere Ausgangsbelastung (HEALTH-49 A zu t0) geht hingegen mit einer geringeren Chance einer Late Response_{TG} einher [OR = 0.43, CI: 0.23 - 0.82, $p = .010$]. Tendenzielle Effekte die das Chancenverhältnis zu Gunsten einer Late Response_{TG} verändern, sind bei den (Ko-)Diagnosen Persönlichkeitsstörung und Adipositas gegeben. Psychische Komorbidität und eine erhöhte Selbstwirksamkeitserwartung zu Behandlungsbeginn senken tendenziell die Chance einer Late Response_{TG}.

Eine *erhöhte Chance für die Zugehörigkeit zu der Non-Response_{TG}-Gruppe* der theoriegeleiteten Verlaufgruppen ergibt sich für Patienten mit längeren Behandlungszeiten [OR = 1.04, CI: 1.01 - 1.08, $p = .022$], längeren Arbeitsunfähigkeitszeiten vor Behandlungsbeginn [OR = 1.06, CI: 1.02 - 1.09, $p = .001$], mit einer Persönlichkeitsstörung [OR = 5.59, CI: 1.65 - 18.87, $p = .022$] und mit ambulanter psychotherapeutischer Vorbehandlung [OR = 2.08, CI: 1.16 - 3.75, $p = .014$]. Dahingegen weisen Patienten mit einer höheren Ausgangsbelastung [OR = 0.06, CI: 0.03 - 0.11, $p < .001$] und mit höheren Werten zur Güte der therapeutischen Beziehung [OR = 0.88, CI: 0.83 - 0.94, $p < .001$] eine geringere Chance auf, in der Gruppe der Non-Response_{TG} zu sein. Eine tendenziell erhöhte Chance einer Non-Response_{TG} zeigt sich für Patienten mit der (Ko-)Diagnose einer somatoformen Störung und für Patienten, die in einer Partnerschaft leben. Bei der Interpretation der absoluten Höhe der OR-Werte für die Behandlungsdauer und die Arbeitsunfähigkeitszeiten muss die Skalierung (in Tagen) berücksichtigt werden.

Die Parameterschätzer der Gruppe n/a werden nicht interpretiert. Dennoch ist zu erwähnen, dass die Klinikvariablen, die sich im Gesamtmodell als signifikant erwiesen haben, in einem signifikanten Zusammenhang mit dieser Gruppe stehen. Die Patienten aus den drei Kliniken weisen demnach bedeutend unterschiedliche Chancen auf, in die Gruppe zu kommen, die durch zu viele Missings gekennzeichnet ist.

Hinsichtlich der zu Fragestellung 2 B postulierten Hypothese zeigt sich auch für die Vorhersage der theoriegeleiteten Verlaufgruppen, dass die theoretisch und empirisch fundierten Prädiktoren einen signifikanten Anteil der Varianz in den theoriegeleiteten Verlaufgruppen erklären. Die *Hypothese 3* wird angenommen. Die Nullhypothese kann abgelehnt werden, da die Modellgüte sich signifikant von Null unterscheidet.

Insgesamt werden 51% der Varianz in den theoriegeleiteten Verlaufgruppen durch die untersuchten Patientenmerkmale erklärt, was der Höhe einer großen Effektstärke (Pseudo- $R^2 \geq .50$; vgl. Tabelle 3, Abschnitt 4.5.4) entspricht.

5.7 Vorhersage des primären Outcomes durch die explorativen Verlaufgruppen (Fragestellung 3 A)

Mit der dritten Fragestellung wurde der Vorhersagewert des Symptomverlaufs auf das längerfristige Behandlungsergebnis untersucht. In einem ersten Schritt werden die Verlaufgruppen im Hinblick auf Unterschiede im katamnestischen Outcome mittels Kovarianzanalyse untersucht. Anschließend wird die Vorhersage des Outcomes durch die Verlaufgruppen unter Kontrolle weiterer relevanter Prädiktoren (Patienten- und Prozessmerkmale) mittels hierarchischer Regressionsanalyse untersucht.

Beide Schritte werden sowohl für die explorativen als auch für die theoriegeleiteten Verlaufgruppen mit dem primären Outcome *Symptombelastung* und dem sekundären Outcome *psychische Aspekte der gesundheitsbezogenen Lebensqualität* durchgeführt. Um eine Vergleichbarkeit der Ergebnisse der explorativen und der theoriegeleiteten Verlaufgruppen zu erzielen, wurde für beide Verlaufgruppeneinteilungen das gleiche Set an Confoundern (zu kontrollierende Prädiktoren) eingeschlossen.

5.7.1 Unterscheiden sich die explorativen Verlaufgruppen im primären Outcome?

Zur Untersuchung von Unterschieden zwischen den explorativen Verlaufgruppen im primären Outcome *katamnestische Symptombelastung* wurden Kovarianzanalysen durchgeführt. Als Kovariate wurde der Ausgangswert des Zielkriteriums, hier die Symptombelastung, eingeschlossen. In Tabelle 33 sind die Ergebnisse der Kovarianzanalyse mit dem Faktor *explorative Verlaufgruppe* und dem katamnestischen HEALTH-49 A Skalenwert als abhängige Variable dargestellt.

Tabelle 33: Ergebnisse der Aufnahme-Katamnese-Vergleiche der *Symptombelastung (HEALTH-49 A)* zwischen den *explorativen Verlaufsgruppen* (ANCOVA; $n = 424$)

	Aufnahme (t0)	Katamnese (t2)	Effekt- stärke <i>d</i>	ANCOVA	
	<i>M</i> / (<i>SD</i>)	<i>M</i> / (<i>SD</i>)	t0 – t2	Faktor: explorative Verlaufsgruppe	Kovariate: Ausgangs- belastung
Gesamtstichprobe ($n = 424$)	1.35 (0.76)	0.93 (0.72)	0.54 ⁺⁺		
Linear Response _{EG} ($n = 314$)	1.05 (0.55)	0.72 (0.53)	0.43 ⁺		
Early Response _{EG} ($n = 41$)	2.25 (0.48)	1.03 (0.74)	1.56 ⁺⁺⁺		
Delayed Response _{EG} ($n =$ 17)	2.78 (0.62)	2.36 (0.71)	0.54 ⁺⁺	$F(4,418) =$ 20.661 $p < .001^{***};$ $\eta^2 = .165^{+++}$	$F(1,418) =$ 130.606 $p < .001^{***};$ $\eta^2 = .238^{+++}$
Non-Response _{EG} ($n = 41$)	1.99 (0.59)	1.73 (0.73)	0.33 ⁺		
Atypical Course ($n = 11$)	2.00 (0.62)	1.30 (0.75)	0.90 ⁺⁺⁺		

M: Mittelwert; *SD*: Standardabweichung; *p*: Signifikanz; η^2 : Effektstärke Eta²; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**) und große (***) Effektstärken interpretiert.

Anmerkung: Effektstärke *d* berechnet aus $(M_{Prä} - M_{Kat}) / SD_{Prä}$ Gesamtstichprobe

Hinsichtlich der Veränderungen des HEALTH-49-A-Skalenwertes zeigt sich, dass für alle Subgruppen des Verlaufs zusammen eine Reduktion der Symptombelastung vom Aufnahme- zum Katamnesezeitpunkt zu verzeichnen ist. Für die einzelnen explorativen Verlaufsgruppen fallen die Effektstärken jedoch in unterschiedlicher Höhe aus. Die Veränderungen in der Linear Response_{EG}- und der Non-Response_{EG}-Gruppe erreichen kleine Effektstärken, in der Delayed-Response_{EG}-Gruppe mittlere und in der Early-Response_{EG}-Gruppe große Effektstärken. Dies spiegelt sich auch in den Ergebnissen der Kovarianzanalyse wider. Der Faktor *Verlaufsgruppe* wird auch unter statistischer Kontrolle der Ausgangsbelastung hoch signifikant und erreicht eine große Effektstärke [$F(4,418) = 20.661$, $p < .001$; $\eta^2 = .165$]. Die Ausgangsbelastung weist erwartungsgemäß einen hoch signifikanten Zusammenhang mit der Symptombelastung zur Katamnese auf [$F(1,418) = 130.606$; $p < .001$; $\eta^2 = .238$]. Die Veränderungen der Symptombelastung von der Aufnahme zur Katamnese sind getrennt für die fünf Verlaufsgruppen in Abbildung 25 dargestellt. In Abbildung 26 sind zusätzlich der Mittelwerte der Symptombelastung zur Entlassung dargestellt.

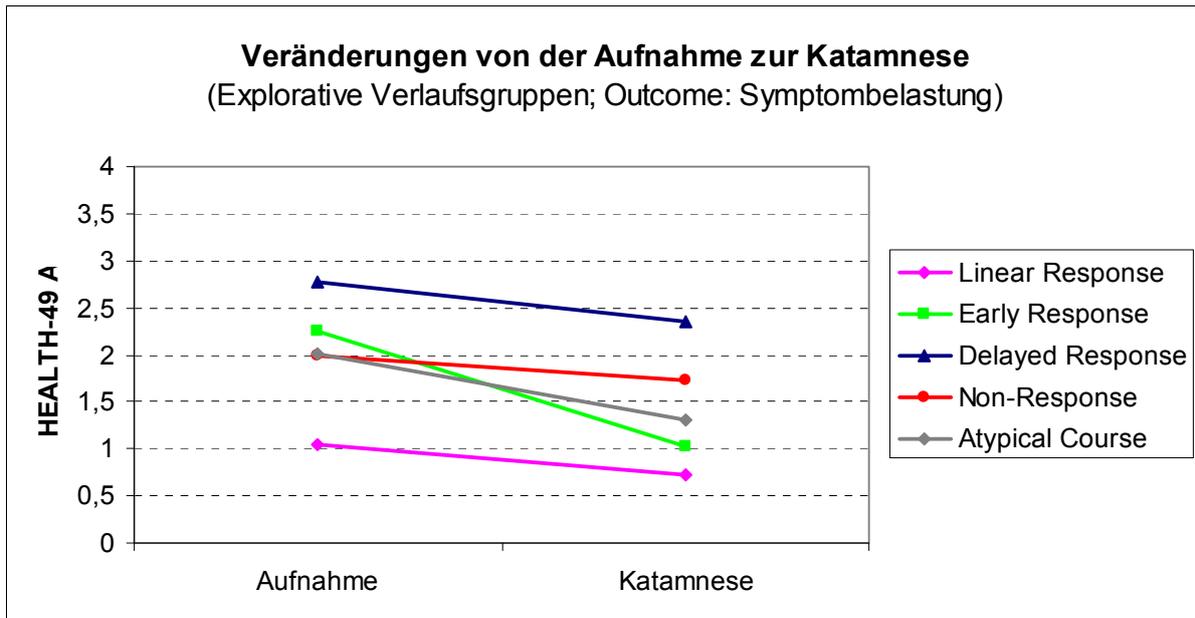


Abbildung 25: HEALTH-49-A-Mittelwerte der explorativen Verlaufsguppen zur Aufnahme und Katamnese

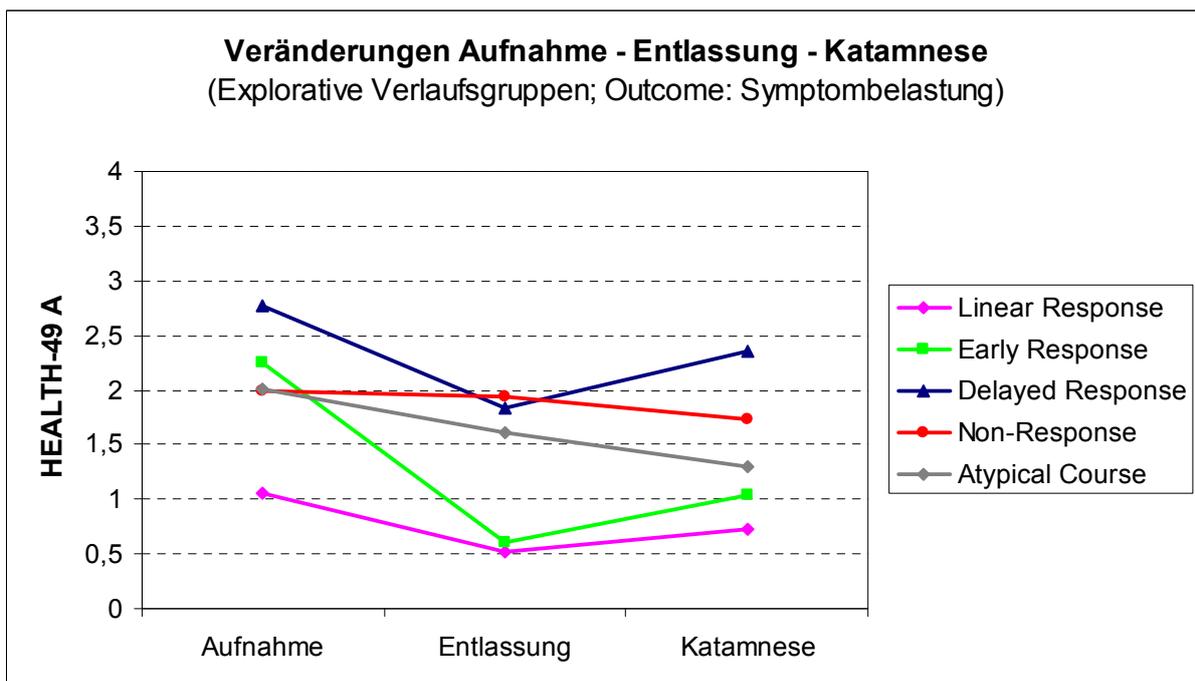


Abbildung 26: HEALTH-49-A-Mittelwerte der explorativen Verlaufsguppen zur Aufnahme, Entlassung und Katamnese

Zur differenzierten Betrachtung der Gruppenunterschiede wurden paarweise Post hoc Vergleiche der explorativen Verlaufsguppen durchgeführt. Dabei wurden die Gruppeneffekte jeweils von dem Effekt der Kovariate Ausgangsbelastung bereinigt. Die Ergebnisse der paarweisen Vergleiche der explorativen Verlaufsguppen sind in Tabelle 34 dargestellt.

Tabelle 34: Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der *Symptombelastung (HEALTH-49 A)* zwischen jeweils zwei *explorativen Verlaufgruppen* (ANCOVAs)

Explorative Verlaufgruppe			Linear Rrsponse	Non- Response	Early Response	Delayed Response
Non-Response	HEALTH t0	p	< .001 ^{***}			
		η^2	.284 ⁺⁺⁺			
	Gruppe	p	< .001 ^{***}			
		η^2	.085 ⁺⁺			
Early Response	HEALTH t0	p	< .001 ^{***}	< .001 ^{***}		
		η^2	.240 ⁺⁺⁺	.315 ⁺⁺⁺		
	Gruppe	p	.003 ^{**}	< .001 ^{***}		
		η^2	.025 ⁺	.345 ⁺⁺⁺		
Delayed Response	HEALTH t0	p	< .001 ^{***}	< .001 ^{***}	.010 [*]	
		η^2	.233 ⁺⁺⁺	.281 ⁺⁺⁺	.114 ⁺	
	Gruppe	p	< .001 ^{***}	.546	< .001 ^{***}	
		η^2	.092 ⁺⁺	.007	.302 ⁺⁺⁺	
Atypical Course	HEALTH t0	p	< .001 ^{***}	< .001 ^{***}	.006 ^{**}	.322
		η^2	.244 ⁺⁺⁺	.342 ⁺⁺⁺	.144 ⁺⁺⁺	.039 ⁺
	Gruppe	p	.409	.038 [*]	.095 [°]	.014 [*]
		η^2	.002	.085 ⁺⁺	.056 ⁺	.218 ⁺⁺⁺

p : Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (****) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Es zeigt sich, dass, bis auf den Vergleich zwischen den Gruppen Non-Response_{EG} und Delayed Response_{EG}, alle Gruppenvergleiche der katamnestic Symptombelastung statistisch hoch signifikant ausfallen. Dabei erreichen die Unterschiede der Gruppenpaare Linear Response_{EG} - Non-Response_{EG}, Linear Response_{EG} - Delayed Response_{EG}, Early Response_{EG} - Non-Response_{EG} und Early Response_{EG} - Delayed Response_{EG} mittlere Effektstärken.

Die in *Hypothese 4* postulierte Annahme, dass sich die *explorativen Verlaufgruppen*, bei statistischer Kontrolle des Ausgangswerts des Ergebnisparameters, bedeutsam im primären Outcome *katamnestic Symptombelastung* unterscheiden, kann angenommen werden. Die Nullhypothese eines nicht signifikanten Unterschieds wird abgelehnt.

Die Unterschiede zwischen den explorativen Verlaufgruppen im primären Outcome *katamnestic Symptombelastung* erreichen die Höhe einer großen Effektstärke.

5.7.2 Auswahl der zu kontrollierenden Prädiktoren

Für die hierarchische Regressionsanalyse wurden weitere Prädiktoren, die einen potentiellen Zusammenhang zum Outcome aufweisen, als zu kontrollierende Variablen berücksichtigt. Um den Pool an literaturbasierten Prädiktoren zu reduzieren, wurde zunächst eine lineare

Regression ausschließlich mit den literaturbasierten Prädiktoren zur Vorhersage des Outcomes durchgeführt. In diese Regressionsanalyse wurden die 23 soziodemographischen und klinischen Patientenmerkmale, die auch für die logistischen Regressionen der Fragestellung 2 untersucht wurden, eingeschlossen (vgl. Tabelle 26, Abschnitt 5.6.3.4, S. 143). Das Verhältnis von Prädiktoren- ($k = 23$) zu Patientenzahl ($n = 375$) liegt mit 1:16 oberhalb des geforderten Verhältnisses von 1:15, um eine ausreichende Power der Analyse zu gewährleisten (Cohen & Cohen, 1983; Stevens, 2009). Die Ergebnisse zeigten für die folgenden 9 der untersuchten Merkmale einen signifikanten Vorhersagewert hinsichtlich der katamnestischen Symptombelastung: Schulabschluss, Erwerbstätigkeit, Depression, Anpassungsstörung, Arbeitsunfähigkeitszeiten, FPTM Leidensdruck, FPTM Hoffnung, HAQ und SWE. Der Anteil der aufgeklärten Varianz lag bei 32%. Aus den Ergebnissen dieser Regression wurden anschließend nur die Variablen in die hierarchische Regressionsanalyse eingeschlossen, die einen signifikanten Vorhersagewert für das Outcome aufwiesen. Die Skala *Leidensdruck* des FPTM wurde zudem aufgrund inhaltlicher Erwägungen aus der Analyse ausgeschlossen: In Fragestellung 2 erwies sich diese Skala als eine Variable, die mit den explorativen Verlaufgruppen bedeutend zusammenhängt und so den Zusammenhang der Verlaufgruppen mit dem Outcome überdecken könnte. Ferner fällt der Zusammenhang zwischen der *Leidensdruck*-Skala und der Ausgangsbelastung (HEALTH-49 A t0) relativ hoch aus ($r = .60$), sodass das Risiko der Multikollinearität bestehen würde (vgl. Abschnitt 5.6.4.1).

Die verbleibenden 8 Patientenmerkmale wurden im ersten Block der hierarchischen Regressionsanalyse untersucht. Als zweiter Block wurden die dummykodierte(n) Klinikvariablen eingefügt. Den dritten Block bildeten die Prozessmerkmale Behandlungsdauer, Dosis und Art der psychotherapienahen Maßnahmen. In den vierten Block wurden die fokussierten Verlaufgruppen eingeschlossen. Die jeweiligen Referenzgruppen der dummykodierte(n) Gruppenvariablen waren, wie schon bei der Durchführung der logistischen Regression, jeweils die größten Gruppen (explorative Verlaufgruppen: Linear Response_{EG}; theoriegeleitete Verlaufgruppen: Early Response_{TG}; vgl. Abschnitt 5.6.4).

Die Ausgangsbelastung, hier der Skalenwert im HEALTH-49 A zu t0, wurde zur Bildung der explorativen Verlaufgruppen einbezogen. Da die Verlaufgruppen demnach über die Veränderung im Verlauf *und* die Ausgangsbelastung definiert sind, würde ein Einschluss des Prädiktors Ausgangsbelastung im ersten Block (Patientenmerkmale) möglicherweise Effekte der Vorhersage durch die Verlaufgruppen überdecken. So wäre durch die Untersuchung der explorativen Verlaufgruppen nur eine Aussage über die *Veränderung* (also die Steigung der Verlaufskurven) möglich, nicht aber über die hier definierten Verlaufgruppen. Um den Varianzanteil, der ausschließlich durch die Ausgangsbelastung erklärt wird, nicht

unberücksichtigt zu lassen, wird diese in einem letzten Block untersucht. Innerhalb jedes Blocks wurden alle Prädiktoren gleichzeitig getestet (Methode Enter).

5.7.3 Vorhersage des primären Outcomes durch die explorativen Verlaufsgruppen

Zur Prüfung der Voraussetzungen der linearen Regression wurde die Kennwerte der Multikollinearität und Autoregression der Residuen berechnet. Die Toleranz der Koeffizienten liegt zwischen 0.38 und 0.89 (*VIF* 1.13 - 2.66) und damit deutlich außerhalb der Grenzen ($T < 0.1$; $VIF > 10$), die auf Multikollinearität hinweisen. Der Durbin-Watson-Koeffizient liegt bei 2.16, die Voraussetzung der unsystematischen Residuen ist somit auch weitgehend erfüllt. In Tabelle 35 sind die Modellparameter der fünf Blöcke der hierarchischen Regressionsanalyse dargestellt.

Tabelle 35: Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Symptombelastung (HEALTH-49 A zu t2)* durch die *explorativen Verlaufsgruppen* ($n = 348$)

Modell	Prädiktorenblock	R^2	Änderungsstatistik	
			inkrementelles R^2	p
1	Patientenmerkmale	.31	.31	< .001***
2	1 + Klinik	.31	< .01	.367
3	2 + Prozessmerkmale	.32	< .01	.219
4	3 + expl. Verlaufsgruppen	.47	.15	< .001***
5	4 + Ausgangsbelastung	.55	.08	< .001***

p : Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (*), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktorenblöcke (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Die Ergebnisse zur Modelgüte zeigen, dass 31% der Varianz in der katamnestischen Symptombelastung durch die in dem ersten Modell getesteten Patientenmerkmale erklärt werden. Die Hinzunahme der Klinikvariablen (2. Modell) erhöht die Varianzaufklärung nicht signifikant, der inkrementelle Anteil liegt unter 1% [$R^2_{\text{inkrementell}} < .01$; $p = .367$]. Ebenso leistet die Hinzunahme der Prozessmerkmale Behandlungsdauer, Dosis und Art psychotherapienaher Maßnahmen keinen bedeutenden Beitrag zur Varianzaufklärung [inkrementelles $R^2_{\text{inkrementell}} < .01$; $p = .219$]. In dem vierten Modell wird der Zuwachs an aufgeklärter Varianz durch die explorativen Verlaufsgruppen getestet und erweist sich mit einem inkrementellen R^2 von .15 als hoch signifikant [$R^2_{\text{inkrementell}} = .15$; $p < .001$]. Das bedeutet, dass zusätzlich 15% der Varianz im primären Outcome durch die explorativen Verlaufsgruppen erklärt werden können, bei Berücksichtigung des Anteils an Varianzaufklärung durch theoretisch fundierte Patienten- und Prozessmerkmale. Die Ausgangsbelastung, hier der Skalenwert des HEALTH-49 A zu t0, erklärt zusätzlich zu allen

zuvor eingeschlossenen Prädiktoren noch einen Anteil von 8% der Varianz im Skalenwert zu t2. Über das Gesamtmodell werden somit 55% der Varianz im primären Outcome aufgeklärt. Die standardisierten Regressionskoeffizienten (Beta-Koeffizienten) mit dem jeweiligen Signifikanzniveau aller getesteten Prädiktoren sind Tabelle 36 zu entnehmen.

Tabelle 36: Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage des *katamnestischen Symptombelastung (HEALTH-49 A zu t2)* durch die *explorativen Verlaufsguppen (n = 348)*

Prädiktor	Beta-Koeffizient	p
Patientenmerkmale		
Schulabschluss hoch	-.109	.007**
Erwerbstätigkeit	-.095	.019*
Anpassungsstörung	-.141	.021*
Depression	-.076	.209
Arbeitsunfähigkeitstage	.055	.181
FPTM Hoffnung	-.045	.274
Therapeutische Beziehung	-.114	.004**
Selbstwirksamkeitserwartung	-.085	.062°
Klinik		
Klinik 1	-.041	.390
Klinik 3	-.037	.470
Prozessmerkmale		
Behandlungsdauer	-.104	.027*
Dosis	.023	.682
Art	.002	.958
Explorative Verlaufsguppen		
Early Response _{EG}	-.117	.011*
Delayed Response _{EG}	.189	< .001***
Non-Response _{EG}	.138	.001***
Atypical Course	.010	.799
Ausgangsbelastung		
HEALTH-49 A zu t0	.449	< .001***

p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**) und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktoren (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Hinsichtlich der Patientenmerkmale zeigt sich, dass die Variablen Schulabschluss, Erwerbstätigkeit, Anpassungsstörung und die therapeutische Beziehung eine signifikante Vorhersage der katamnestischen Symptombelastung erlauben. Dabei hat die Diagnose einer Anpassungsstörung mit einem Beta von $-.144$ [$p = .021$] den höchsten Vorhersagewert unter den Patientenmerkmalen. Hinsichtlich der Richtung der Vorhersage lässt sich sagen, dass Patienten mit hohem Schulabschluss, Patienten, die erwerbstätig sind und Patienten mit einer Anpassungsstörung eine niedrigere katamnestische Symptombelastung aufweisen. Je

besser die Güte der therapeutischen Beziehung in der frühen Behandlungsphase eingeschätzt wird (hoher Skalenwert), umso niedriger ist die Symptombelastung 6 Monate nach der Behandlung. Die Selbstwirksamkeitserwartung zu Behandlungsbeginn hat einen negativ gerichteten tendenziellen Vorhersagewert auf die katamnestische Symptombelastung [Beta = -.085; $p = .062$]. Die behandelnden Kliniken haben keinen bedeutenden Vorhersagewert, ebenso wie die Prozessvariablen Dosis und Art psychotherapienaher Behandlungsmaßnahmen. Allerdings erweist sich die Behandlungsdauer als signifikanter Prädiktor: Längere Behandlungszeiten gehen demnach mit einer geringeren Symptombelastung einher [Beta = -.104; $p = .027$].

Hinsichtlich der im Fokus stehenden explorativen Verlaufgruppen erweisen sich drei der vier eingeschlossenen Gruppen (Linear Response_{EG} als Referenzgruppe) als signifikante Prädiktoren. Die Zugehörigkeit zur Early-Response_{EG}-Gruppe sagt eine geringere [Beta = -.117; $p = .011$], die Gruppen Delayed Response_{EG} [Beta = .189; $p < .001$] und Non-Response_{EG} [Beta = .138; $p = .001$] hingegen eine höhere Symptombelastung zur Katamnese vorher. Dabei leistet die Delayed Response_{EG} den größten Erklärungsbeitrag der explorativen Verlaufgruppen. Der Ausgangswert der Symptombelastung sagt ebenfalls signifikant den katamnestischen Wert der Symptombelastung vorher. Patienten, die zur Aufnahme höhere Werten im HEALTH-49 A aufweisen, haben auch zur Katamnese höhere Skalenwerte [Beta = .449; $p < .001$]. Den höchsten Vorhersagewert gegenüber den anderen Prädiktoren leistet die Ausgangsbelastung.

Für das untersuchte Modell wurden der Modellparameter $R^2_{\text{inkrementell}}$ in die Effektgröße f^2 umgerechnet³¹. Ein inkrementelles R^2 von .15 erreicht demnach mit einem f^2 von 0.33 eine annähernd große Effektstärke.

Die in *Hypothese 5* postulierte Annahme, dass durch die explorativen Verlaufgruppen ein signifikanter inkrementeller Anteil der Varianz im primären Outcome *katamnestische Symptombelastung* erklärt wird, kann damit angenommen werden. Die Nullhypothese wird abgelehnt, da der inkrementelle Varianzanteil sich bedeutend von Null unterscheidet.

Der durch die explorativen Verlaufgruppen erklärte Varianzanteil im primären Outcome *katamnestische Symptombelastung* erreicht die Höhe einer annähernd großen Effektstärke.

Aufgrund des hohen Vorhersagewertes des Prädiktors *Ausgangsbelastung* wurde eine zweite Regression durchgeführt, in der die Ausgangsbelastung in den ersten Block der Patientenmerkmale aufgenommen wurde³². Durch dieses Vorgehen soll ein

³¹ Die Formel zur Berechnung der Effektgröße f^2 lautet: $f^2 = R^2_{\text{inkrementell}} / (1 - R^2_{\text{Gesamt}})$.

³² Die inhaltliche Begründung für das Vorgehen in den oben dargestellten Hauptanalysen ist Abschnitt 5.7.2 zu entnehmen.

Informationsgewinn über den Zuwachs an Varianzaufklärung durch die Verläufe, bei Konstanthaltung der Ausgangsbelastung, erzielt werden. Die Ergebnisse zeigen, dass durch die Patientenmerkmale (inkl. Ausgangsbelastung) 48% der Varianz im Outcome erklärt werden. Klinik- und Prozessvariablen erbringen keinen signifikanten Zuwachs, die explorativen Verlaufsgruppen, bzw. genauer die Verlaufsänderungen der explorativen Gruppen, klären jedoch immer noch einen signifikanten Anteil von 6% der Varianz auf [$R^2_{\text{inkrementell}} = .06$; $p < .001^{***}$]. Über das Gesamtmodell werden 55% der Varianz im Outcome erklärt. Der inkrementelle Zuwachs durch die explorativen Verlaufsgruppen entspricht hier einer kleinen Effektstärke [$f^2 = 0.13$].

5.8 Vorhersage des primären Outcomes durch die theoriegeleiteten Verlaufsgruppen (Fragestellung 3 B)

5.8.1 Unterscheiden sich die theoriegeleiteten Verlaufsgruppen im primären Outcome?

Die Mittelwertsvergleiche und Effektstärken der Veränderung der Symptombelastung zwischen Aufnahme und Katamnese für die vier theoriegeleiteten Verlaufsgruppen sowie die Ergebnisse der Kovarianzanalyse (Kovariate: HEALTH-49 A zu t0) sind in Tabelle 37 dargestellt.

Tabelle 37: Ergebnisse der Aufnahme-Katamnese-Vergleiche der *Symptombelastung* (HEALTH-49 A) zwischen den *theoriegeleiteten Verlaufsgruppen* (ANCOVA; $n = 424$)

	Aufnahme (t0)	Katamnese (t2)	Effekt- stärke <i>d</i>	ANCOVA	
	<i>M</i> / (<i>SD</i>)	<i>M</i> / (<i>SD</i>)	t0 – t2	Faktor: theoriegeleitete Verlaufsgruppe	Kovariate: Ausgangs- belastung
Gesamtstichprobe ($n = 424$)	1.35 (0.76)	0.93 (0.72)	0.54 ⁺⁺		
Early Response _{TG} ($n = 207$)	1.71 (0.68)	1.02 (0.71)	0.89 ⁺⁺⁺		
Late Response _{TG} ($n = 40$)	1.37 (0.56)	0.91 (0.73)	0.60 ⁺⁺	$F(3,419) =$ 11.734	$F(1,419) =$ 329.725
Non-Response _{TG} ($n = 112$)	0.83 (0.60)	0.78 (0.72)	0.06	$p < .001^{***};$ $\eta^2 = .078^{++}$	$p < .001^{***};$ $\eta^2 = .440^{+++}$
n/a ($n = 65$)	1.13 (0.76)	0.92 (0.72)	0.27 ⁺		

M: Mittelwert; *SD*: Standardabweichung; *p*: Signifikanz; η^2 : Effektstärke Eta²; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (°), mittlere (**), und große (***) Effektstärken interpretiert.

Anmerkung: Effektstärke *d* berechnet aus $(M_{\text{Prä}} - M_{\text{Kat}}) / SD_{\text{Prä}}$ Gesamtstichprobe

Sowohl in der Gesamtgruppe als auch in den Subgruppen Early Response_{TG} und Late Response_{TG} verbessert sich die Symptombelastung von der Aufnahme zur Katamnese mit

mittlerer bis großer Effektstärke (Early Response_{TG}). Die Non-Response_{TG}-Gruppe zeigt fast keine Veränderung der Symptombelastung. Die Ergebnisse der Kovarianzanalyse zeigen, dass sich die theoriegeleiteten Verlaufgruppen auch unter statistischer Kontrolle der Ausgangsbelastung signifikant im Outcome unterscheiden [$F(3,419) = 11.734$; $p < .001$; $\eta^2 = .078$]. Das bedeutet, dass auch bei Konstanzhaltung der Ausgangsbelastung ein bedeutender Unterschied zwischen den theoriegeleiteten Verlaufgruppen zu verzeichnen ist. Anders formuliert wird der Zusammenhang zwischen den theoriegeleiteten Verlaufgruppen und der katamnestischen Symptombelastung signifikant und erreicht das Ausmaß einer mittleren Effektstärke. Die Kovariante Ausgangsbelastung weist ebenfalls einen signifikanten Effekt auf [$F(1,419) = 329.725$; $p < .001^{***}$; $\eta^2 = .440$]. Die Veränderungen der Symptombelastung in den theoriegeleiteten Verlaufgruppen sind in Abbildung 27 und Abbildung 28 dargestellt.

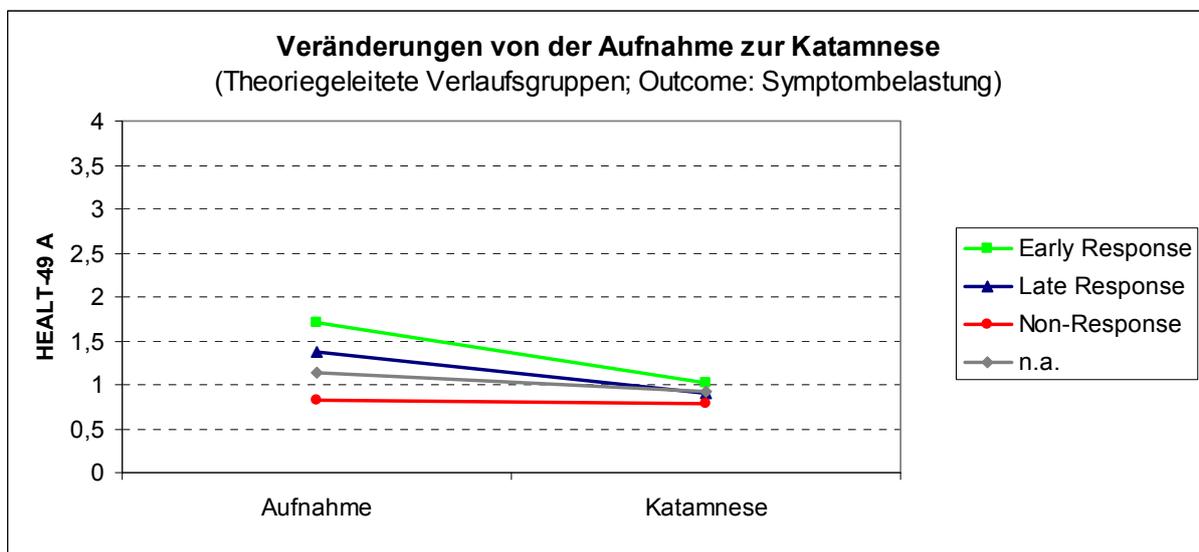


Abbildung 27: HEALTH-49-A-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme und Katamnese

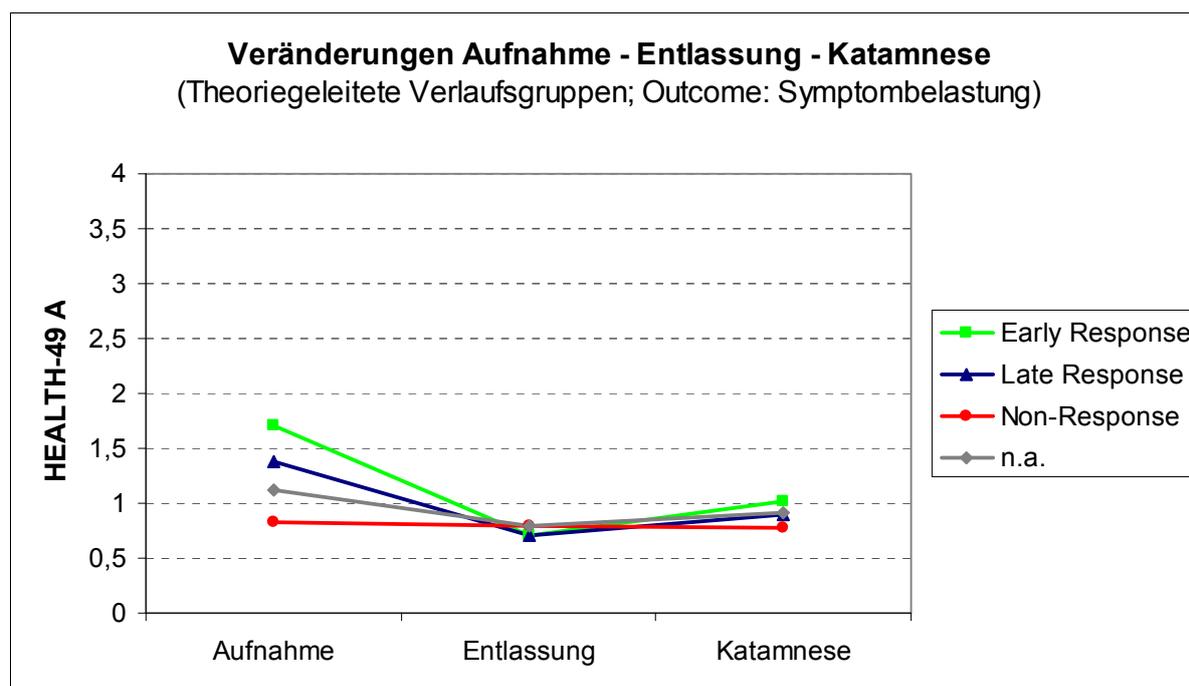


Abbildung 28: HEALTH-49-A-Mittelwerte der theoriegeleiteten Verlaufsgruppen zur Aufnahme, Entlassung und Katamnese

In den Abbildungen wird deutlich, dass sich die Differenzierung der Symptombelastung zwischen den theoriegeleiteten Verlaufsgruppen vom Beginn der Behandlung bis zur Katamnese deutlich verringert. Zur differenzierten Betrachtung der einzelnen Verlaufsgruppen wurden paarweise Vergleiche mit der Ausgangsbelastung als Kovariate durchgeführt. Die Ergebnisse sind in Tabelle 38 dargestellt.

Tabelle 38: Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der *Symptombelastung* (HEALTH-49 A) zwischen jeweils zwei *theoriegeleiteten Verlaufsgruppen* (ANCOVAs)

Theoriegeleitete Verlaufsgruppe			Non-Response	Early Response	Late Response
Early Response	HEALTH t0	p	< .001***		
		η^2	.394 ⁺⁺⁺		
	Gruppe	p	< .001***		
		η^2	.067 ⁺⁺		
Late Response	HEALTH t0	p	< .001***	< .001***	
		η^2	.562 ⁺⁺⁺	.360 ⁺⁺⁺	
	Gruppe	p	< .001***	.284	
		η^2	.094 ⁺	.005	
n/a	HEALTH t0	p	< .001***	< .001***	< .001***
		η^2	.564 ⁺⁺⁺	.392 ⁺⁺⁺	.590 ⁺⁺⁺
	Gruppe	p	.162	.001***	.027*
		η^2	.011 ⁺	.038 ⁺	.047 ⁺

p : Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Die Non-Response_{TG}-Gruppe unterscheidet sich, bei Konstanthaltung der Ausgangsbelastung, signifikant von der Early-Response_{TG}-Gruppe [$p < .001$; $\eta^2 = .067$] und der Late-Response_{TG}-Gruppe [$p < .001$; $\eta^2 = .094$] hinsichtlich des Outcomes. Dahingegen ist der Unterschied zwischen der Early-Response_{TG}- und der Late-Response_{TG}-Gruppe nicht signifikant [$p = .284$; $\eta^2 = .005$]. Entsprechend der Ergebnisse der Kovarianzanalyse über alle Gruppen erweist sich die Ausgangsbelastung für alle paarweisen Vergleiche als hoch signifikant.

Die in *Hypothese 6* postulierte Annahme, dass sich die theoriegeleiteten Verlaufgruppen, bei statistischer Kontrolle des Ausgangswertes des Ergebnisparameters, bedeutsam im primären Outcome *katamnestische Symptombelastung* unterscheiden, kann angenommen werden. Die Nullhypothese eines nicht signifikanten Unterschieds wird abgelehnt.

Die Unterschiede zwischen den theoriegeleiteten Verlaufgruppen im primären Outcome *katamnestische Symptombelastung* erreichen die Höhe einer mittleren Effektstärke.

5.8.2 Auswahl der zu kontrollierenden Prädiktorvariablen

Zur Vorhersage des Outcomes über die theoriegeleiteten Verlaufgruppen wurden dieselben Prädiktorvariablen (Patienten- und Prozessmerkmale) in die hierarchische Regression aufgenommen, die auch zur Vorhersage über die explorativen Verlaufgruppen eingeschlossen wurden, (vgl. 5.7.2). Demnach wurden 8 Patientenmerkmale im ersten Block, 2 Klinikvariablen (dummykodiert) im zweiten Block, 3 Prozessmerkmale im dritten Block und die 3 im Fokus stehenden theoriegeleiteten Verlaufgruppen im vierten Block untersucht. In den fünften Block wurde die Ausgangsbelastung aufgenommen³³.

5.8.3 Vorhersage des primären Outcomes durch die theoriegeleiteten Verlaufgruppen

Die Voraussetzungen der linearen Regression sind mit Toleranzwerten von 0.38 – 0.87 (*VIF*: 1.31 – 2.65) und einem Durbin-Watson-Koeffizienten von 2.20 erfüllt. In Tabelle 39 sind die Kennwerte der Modellgüte der fünf getesteten Modelle aufgeführt.

³³ In die Bildung der theoriegeleiteten Verlaufgruppen wurde die Ausgangsbelastung zwar nur indirekt aufgenommen (vgl. Abschnitt 4.3.4), um die Ergebnisse der theoriegeleiteten Verlaufgruppen mit den Ergebnissen der explorativen Verlaufgruppen vergleichen zu können, wird die Ausgangsbelastung hier auch im letzten (und nicht im ersten Block) untersucht.

Tabelle 39: Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Symptombelastung (HEALTH-49 A zu t2)* durch die *theoriegeleiteten Verlaufsguppen (n = 348)*

Modell	Prädiktorenblock	R^2	Änderungsstatistik	
			inkrementelles R^2	p
1	Patientenmerkmale	.31	.31	< .001***
2	1 + Klinik	.31	< .01	.434
3	2 + Prozessmerkmale	.32	< .01	.216
4	3 + theor. Verlaufsguppen	.33	< .01	.336
5	4 + Ausgangsbelastung	.50	.17	< .001***

p : Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktorenblöcke (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Die ersten drei getesteten Modelle entsprechen denen, die schon für die explorativen Verlaufsguppen untersucht wurden. Die Patientenmerkmale erklären 31% der Varianz in der katamnestischen Symptombelastung. Durch die Hinzunahme der Klinikvariablen, ebenso wie der Prozessmerkmale Behandlungsdauer, Dosis und Art psychotherapeutischer Maßnahmen, kommt es zu keiner signifikanten zusätzlich Varianzaufklärung (inkrementelles R^2 für beide Blöcke jeweils < 1%). Im vierten Modell, in dem die fokussierten theoriegeleiteten Verlaufsguppen getestet wurden, konnte ebenfalls kein signifikanter Zuwachs an Varianzaufklärung verzeichnet werden [$R^2_{\text{inkrementell}} = .007$; $p = .336$]. Durch die Zugehörigkeit zu den theoriegeleiteten Verlaufsguppen kann, unter Berücksichtigung der untersuchten Patienten- und Prozessmerkmale, demnach weniger als 1% zusätzlicher Varianz im Outcome erklärt werden. Die im fünften Modell getestete Hinzunahme der Ausgangsbelastung erhöht die Varianzaufklärung um 17%. Insgesamt werden 50% der Varianz im Outcome über alle untersuchten Prädiktoren erklärt. In Tabelle 40 sind die Parameter der untersuchten Prädiktoren dargestellt.

Tabelle 40: Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Symptombelastung (HEALTH-49 A zu t2)* durch die *theoriegeleiteten Verlaufsgruppen (n = 348)*

Prädiktor	Beta-Koeffizient	p
Patientenmerkmale		
Schulabschluss hoch	-.134	.001***
Erwerbstätigkeit	-.101	.015*
Anpassungsstörung	-.138	.030*
Depression	-.083	.109
Arbeitsunfähigkeitstage	.072	.095°
FPTM Hoffnung	-.062	.105
Therapeutische Beziehung	-.131	.002**
Selbstwirksamkeitserwartung	-.068	.153
Klinik		
Klinik 1	-.042	.404
Klinik 3	-.052	.323
Prozessmerkmale		
Behandlungsdauer	-.091	.067°
Dosis	.009	.879
Art	-.019	.664
Theoriegeleitete Verlaufsgruppen		
Late Response _{TG}	.059	.161
Non-Response _{TG}	.155	.002**
n/a	.082	.069°
Ausgangsbelastung		
HEALTH-49 A zu t0	.583	< .001***

p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**) und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktoren (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

In den Koeffizienten der Patienten- und Prozessmerkmale treten nur marginale Unterschiede zu denen der Regressionsanalyse zu den explorativen Verlaufsgruppen auf. Die Vorhersagekraft der jeweiligen Prädiktoren ist in der gleichen Weise wie bei der Untersuchung der explorativen Verlaufsgruppen zu interpretieren: Patienten, die einen hohen Schulabschluss haben, erwerbstätig sind und Patienten die eine Anpassungsstörung als (Ko-) Diagnose aufweisen, gehen mit geringeren Werten der katamnestischen Symptombelastung einher. Ebenso sagt eine bessere therapeutische Beziehung eine geringere Symptomlast vorher. Die Klinikvariablen und die Prozessmerkmale erweisen sich als nicht signifikante Prädiktoren. Für die Behandlungsdauer ist lediglich ein tendenzieller negativer Zusammenhang zur katamnestischen Symptombelastung zu verzeichnen.

In der differenzierten Betrachtung der Beta-Koeffizienten der theoriegeleiteten Verlaufsgruppen zeigt sich, dass die Gruppe Non-Response_{TG} einen signifikanten

Vorhersagewert des Outcomes leistet [Beta = .155; $p = .002$]. Patienten der Gruppe Non-Response_{TG} weisen eine höhere Symptombelastung zur Katamnese auf (Referenzgruppe: Early Response_{TG}). Der Prädiktor Late Response_{TG} erweist sich hingegen als nicht signifikant [Beta = .059; $p = .161$]. Die Ausgangsbelastung erbringt hier, wie bei der Untersuchung der explorativen Verlaufsgruppen auch, den größten Vorhersagewert für das Outcome [Beta = .583; $p < .001$].

Für das untersuchte Modell wurden der Modellparameter $R^2_{\text{inkrementell}}$ in die Effektgröße f^2 umgerechnet. Ein inkrementelles R^2 von .007 erreicht demnach mit $f^2 = 0.014$ nicht das Ausmaß einer kleinen Effektstärke.

Die in *Hypothese 7* postulierte Annahme, dass durch die theoriegeleiteten Verlaufsgruppen ein signifikanter inkrementeller Anteil der Varianz im primären Outcome erklärt wird, kann *nicht angenommen* werden. Die Nullhypothese wird beibehalten, da der inkrementelle Varianzanteil sich nicht bedeutend von Null unterscheidet.

Ebenso wie in Abschnitt 5.7.3 werden abschließend die Ergebnisse der Regressionsanalyse zum Vorhersagewert der theoriegeleiteten Verlaufsgruppen bei *vorheriger* Berücksichtigung des Prädiktors Ausgangsbelastung (Aufnahme in den ersten Block) dargestellt. Durch die Patientenmerkmale (inkl. Ausgangsbelastung) werden 48% der Varianz im Outcome erklärt. Die Aufnahme von Klinik- und Prozessvariablen erbringt keinen zusätzlichen signifikanten Anteil an Varianzaufklärung. Durch die Hinzunahme der theoriegeleiteten Verlaufsgruppen erhöht sich die erklärte Varianz um 2% [$R^2_{\text{inkrementell}} = .016$; $p = .015$; $f^2 = 0.032$] signifikant mit kleiner Effektstärke auf 50%.

5.9 Vorhersage des sekundären Outcomes durch die explorativen Verlaufsgruppen (Fragestellung 3 C)

5.9.1 Unterscheiden sich die explorativen Verlaufsgruppen im sekundären Outcome?

Zur Unterschiedstestung des sekundären Outcomes *Psychische Aspekte der gesundheitsbezogenen Lebensqualität* wurden Kovarianzanalysen mit den Verlaufsgruppen als Faktor und dem Aushangswert der psychischen Summenskala des SF-8 (im Folgenden mit *SF-8 Psych.* abgekürzt) als Kovariate durchgeführt. Die Skalenmittelwerte des SF-8 Psych. zur Aufnahme und zur Katamnese sind für die explorativen Verlaufsgruppen mit den erzielten Effektstärken und den Ergebnissen der Kovarianzanalyse in Tabelle 41 dargestellt.

Tabelle 41: Ergebnisse der Aufnahme-Katamnese-Vergleiche der *gesundheitsbezogenen Lebensqualität (SF-8 Psych.)* zwischen den *explorativen Verlaufsgruppen* (ANCOVA; $n = 412$)

	Aufnahme	Katamnese	Effekt-	ANCOVA	
	(t0)	(t2)	stärke	Faktor:	Kovariate:
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>d</i>	Explorative	Ausgangs-
			t0 – t2	Verlaufsguppe	belastung
Gesamtstichprobe ($n = 412$)	32.7 (12.1)	40.3 (12.5)	-0.63 ⁺⁺		
Linear Response _{EG} ($n = 306$)	35.4 (12.1)	42.5 (12.0)	-0.59 ⁺⁺		
Early Response _{EG} ($n = 40$)	24.7 (7.8)	38.1 (13.0)	-1.10 ⁺⁺⁺		
Delayed Response _{EG} ($n = 15$)	23.4 (8.4)	28.8 (9.3)	-0.44 ⁺	$F(4,406) = 7.330$ $p < .001^{***}$; $\eta^2 = .067^{++}$	$F(1,406) = 0.015$ $p < .001^{***}$; $\eta^2 = .123^{+++}$
Non-Response _{EG} ($n = 41$)	26.4 (7.9)	30.6 (8.9)	-0.34 ⁺		
Atypical Course ($n = 10$)	22.6 (11.8)	38.7 (14.7)	-1.33 ⁺⁺⁺		

M: Mittelwert; *SD*: Standardabweichung; *p*: Signifikanz; η^2 : Effektstärke Eta²; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 10%- ([°]), dem 5%- (^{*}), dem 1%- (^{**}) und dem 0.1%- Niveau (^{***}) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Anmerkung: Effektstärke *d* berechnet aus $(M_{Prä} - M_{Kat}) / SD_{Prä}$ Gesamtstichprobe

Es wird deutlich, dass sowohl in der Gesamtstichprobe, als auch in den Verlaufssubgruppen von Behandlungsbeginn zur Katamnese eine Steigerung der Lebensqualität auftrat. Dabei erreichten die Veränderungen der fünf Subgruppen unterschiedlich große Effektstärken. Während in der Non-Response_{EG}- und der Delayed-Response_{EG}-Gruppe kleine Effektstärken auftreten, werden in der Linear-Response_{EG}-Gruppe, in der der Ausgangswert am höchsten liegt, mittlere Effektstärken erreicht. Die Early-Response_{EG}-Gruppe weist bei einem relativ niedrigen Mittelwert der Lebensqualität zu Behandlungsbeginn große Effektstärke der Veränderung zur Katamnese auf. Der Effekt der explorativen Verlaufsguppen auf die katamnestiche Lebensqualität wird auch unter statistischer Kontrolle der Ausgangsbelastung hoch signifikant [$F(4,406) = 7.330$; $p < .001^{***}$; $\eta^2 = .067^{++}$]. Der größere Teil an Varianzaufklärung in der katamnestiche Lebensqualität erfolgt dennoch über den Ausgangswert der Lebensqualität (SF-8 Psych. zu t0) [$F(1,406) = 0.015$; $p < .001^{***}$; $\eta^2 = .123$]. In Abbildung 29 und Abbildung 30 sind die Veränderungen der Lebensqualität über die zwei bzw. drei Messzeitpunkte für die explorativen Verlaufsguppen dargestellt.

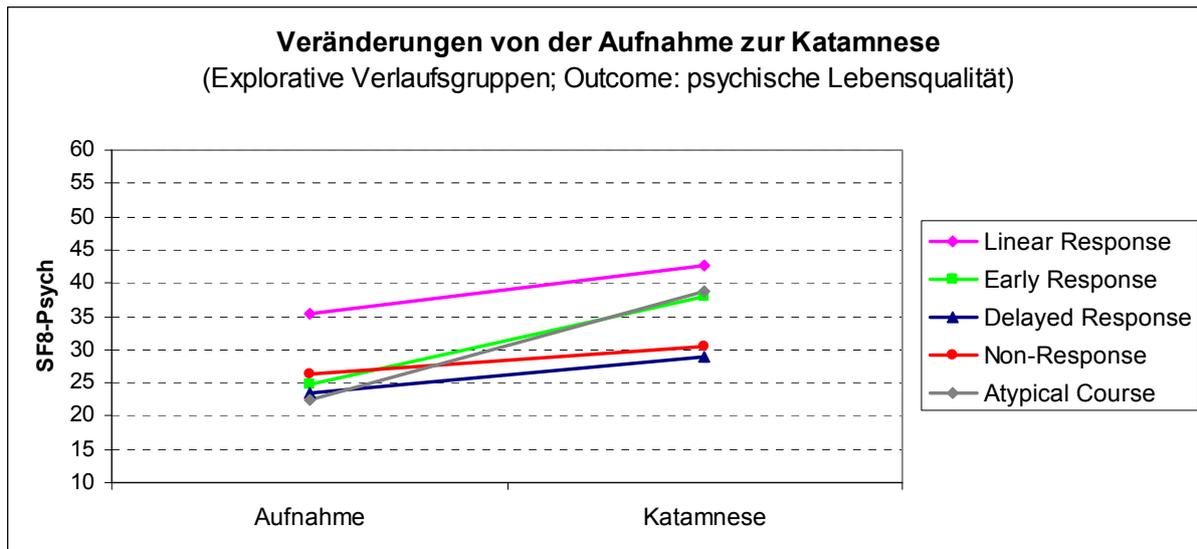


Abbildung 29: SF-8-Psych.-Mittelwerte der explorativen Verlaufsguppen zur Aufnahme und Katamnese

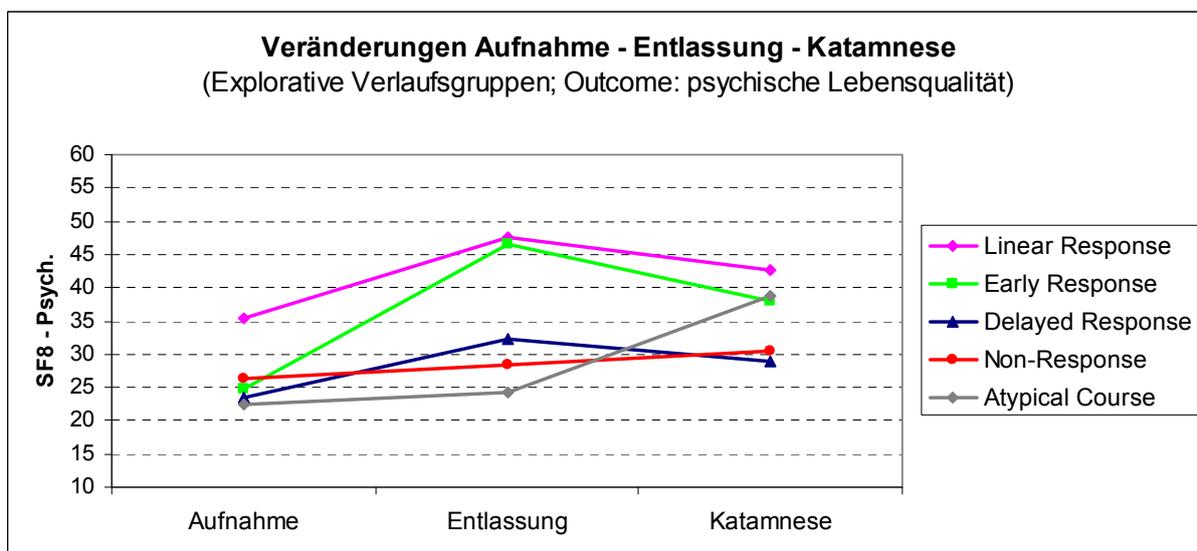


Abbildung 30: SF-8-Psych.-Mittelwerte der explorativen Verlaufsguppen zur Aufnahme, Entlassung und Katamnese

Zur differenzierten Untersuchung des Zusammenhangs der explorativen Verlaufsguppen mit dem sekundären Outcome wurden paarweise Vergleiche zwischen den explorativen Verlaufsguppen unter Kontrolle des Ausgangswertes der Lebensqualität durchgeführt. Tabelle 42 enthält die Ergebnisse der paarweisen Kovarianzanalysen.

Tabelle 42: Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der *gesundheitsbezogenen Lebensqualität (SF-8 Psych.)* zwischen jeweils zwei *explorativen Verlaufsgruppen* (ANCOVAs)

Explorative Verlaufsgruppe			Linear Response	Non-Response	Early Response	Delayed Response
Non-Response	SF-8 t0	p	< .001***			
		η^2	.165+++			
	Gruppe	p	< .001***			
		η^2	.057 ⁺			
Early Response	SF-8 t0	p	< .001***	.001***		
		η^2	.139+++	.126+++		
	Gruppe	p	.869	.001***		
		η^2	< .001	.142+++		
Delayed Response	SF-8 t0	p	< .001***	.010*	.613	
		η^2	.141+++	.119+++	.005	
	Gruppe	p	.002**	.822	.017*	
		η^2	.030 ⁺	.001	.104+++	
Atypical Course	SF-8 t0	p	< .001***	.049*	.772	.121
		η^2	.136+++	.078 ⁺⁺	.002	.106
	Gruppe	p	.800	.010*	.868	.051
		η^2	< .001	.129+++	.001	.163+++

p : Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- ([°]), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine ([°]), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Die katamnestiche Lebensqualität in der Linear-Response_{EG}-Gruppe unterscheidet sich signifikant vom Outcome der Non-Response_{EG}- und der Delayed-Response_{EG}-Gruppe, nicht jedoch von der katamnestiche Lebensqualität der Early-Response_{EG}-Gruppe. Auch die Early-Response_{EG}-Gruppe unterscheidet sich signifikant von der Non-Response_{EG}- und der Delayed-Response_{EG}-Gruppe. Non-Response_{EG} und Delayed Response_{EG} unterscheiden sich nicht bedeutend hinsichtlich des sekundären Outcomes.

Die in *Hypothese 8* postulierte Annahme, dass sich die explorativen Verlaufsgruppen, bei statistischer Kontrolle des Ausgangswerts des Ergebnisparameters, bedeutsam im sekundären Outcome *katamnestiche erfasste gesundheitsbezogene Lebensqualität (psychische Aspekte)* unterscheiden, kann angenommen werden. Die Nullhypothese eines nicht signifikanten Unterschieds wird abgelehnt.

Die Unterschiede zwischen den explorativen Verlaufsgruppen im sekundären Outcome *katamnestiche erfasste gesundheitsbezogene Lebensqualität (psychische Aspekte)* erreichen die Höhe einer mittleren Effektstärke.

5.9.2 Auswahl der zu kontrollierenden Prädiktorvariablen

Für die Vorhersage des sekundären Outcomes *psychische Aspekte der gesundheitsbezogenen Lebensqualität* wurde die gleiche Vorgehensweise wie bei der Vorhersage des primären Outcomes gewählt. In die vorangehende lineare Regressionsanalyse zur Vorhersage des Outcomes ausschließlich aus Patientenmerkmalen wurde hier der Aufnahmewert des Outcomes, also der SF-8-Psych.-Skalenwert zu t0, mit aufgenommen. Die Skala SF-8 Psych. ist nicht in die Verlaufsgruppenbildung mit eingegangen und kann daher als relevante Prädiktorvariable eingeschlossen werden. Darüber hinaus wurde anders als bei der Vorhersage des primären Outcomes die FPTM-Skala *Leidensdruck* in die Analysen einbezogen (vgl. Abschnitt 5.7.2), da diese Skala inhaltlich weniger nah an dem sekundären als an dem primären Outcome anzusiedeln ist, und mit dem Einschluss des Merkmals ein höherer Informationsgewinn erreicht werden kann. Problematisch bleibt jedoch der in Fragestellung 2 gefundene Zusammenhang der Skala *Leidensdruck* mit den explorativen Verlaufsgruppen, der den Zusammenhang der Verlaufsgruppen mit dem Outcome überdecken könnte. Dies muss bei der Diskussion der Ergebnisse berücksichtigt werden. Alle anderen eingeschlossenen Patientenmerkmale entsprechen denen, die für die Vorhersage des primären Outcomes verwendet wurden. Die Ergebnisse zeigen bei einer Gesamtstichprobe von $n = 365$ eine Modellgüte von $R^2 = .33$, in dem die Prädiktoren Schulabschluss, Erwerbstätigkeit, FPTM Leidensdruck, Therapeutische Beziehung und der Ausgangswert im SF-8 Psych. signifikant sind. In der hierarchischen Regressionsanalyse wurden im ersten Block diese 5 Patientenmerkmale, im zweiten Block die Klinikvariablen, im dritten Block die 3 Prozessvariablen (Behandlungsdauer, Dosis und Art psychotherapienaher Maßnahmen) und im vierten Block die fokussierten explorativen Verlaufsgruppen untersucht.

Die Symptombelastung zu Behandlungsbeginn wird nicht als Prädiktor der katamnestischen gesundheitsbezogenen Lebensqualität untersucht, da diese Teil der Verlaufsgruppenbildung war und demnach den (potentiellen) Vorhersagewert der Verlaufsgruppen überdecken kann (vgl. Abschnitt 5.7.2).

5.9.3 Vorhersage des sekundären Outcomes durch die explorativen Verlaufsgruppen

Die Prüfung der Voraussetzungen der multiplen Regressionsanalyse weist mit Toleranzwerten von 0.43 - 0.93 (*VIF*: 1.08 - 2.35) und einem Durbin-Watson-Koeffizienten von 2.25 weder auf Multikollinearität noch auf eine Autokorrelation der Residuen hin. In Tabelle 43 sind die Ergebnisse zur Modellgüte der untersuchten Modelle dargestellt.

Tabelle 43: Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestic Lebensqualität (SF-8 Psych. zu t2)* durch die *explorativen Verlaufsgruppen (n = 362)*

Modell	Prädiktorenblock	R^2	Änderungsstatistik	
			inkrementelles R^2	p
1	Patientenmerkmale	.30	.30	< .001***
2	1 + Klinik	.30	< .01	.325
3	2 + Prozessmerkmale	.31	< .01	.550
4	3 + explorative Verlaufsgruppen	.33	.02	.031*

p : Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktorenblöcke (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Die fünf untersuchten Patientenmerkmale klären einen signifikanten Anteil von 30% der Varianz in der katamnestic Lebensqualität auf. Die Hinzunahme der Klinikvariablen im 2. Modell sowie der Einschluss der Prozessmerkmale im 3. Modell erbringen keinen signifikanten zusätzlichen Beitrag zur Varianzaufklärung. Das inkrementelle R^2 liegt für beide Blöcke unter einem Prozent. Die Hinzunahme der im Fokus stehenden explorativen Verlaufsgruppen trägt einen signifikanten inkrementellen Anteil von 2% erklärter Varianz bei. Über alle in das Gesamtmodell eingeschlossenen Prädiktoren werden entsprechend 33% der Varianz in der katamnestic Lebensqualität erklärt. Tabelle 44 enthält die Parameter der untersuchten Prädiktoren.

Tabelle 44: Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Lebensqualität (SF-8 Psych. zu t2)* durch die *explorativen Verlaufsgruppen (n = 362)*

Prädiktor	Beta-Koeffizient	p
Patientenmerkmale		
Schulabschluss hoch	.113	.016*
Erwerbstätigkeit	.166	< .001***
FPTM Leidensdruck	-.223	< .001***
Therapeutische Beziehung	.124	.007**
Ausgangswert SF-8 Psych.	.228	< .001***
Klinik		
Klinik 1	.074	.178
Klinik 3	-.075	.199
Prozessmerkmale		
Behandlungsdauer	.078	.164
Dosis	-.083	.218
Art	.007	.894
Explorative Verlaufsgruppen		
Early Response _{EG}	-.008	.868
Delayed Response _{EG}	-.082	.084°
Non-Response _{EG}	-.120	.014*
Atypical Course	.055	.227

p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktoren (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Alle fünf untersuchten Patientenmerkmale weisen einen signifikanten Vorhersagewert der katamnestischen Lebensqualität auf. Patienten mit hohem Schulabschluss und Patienten, die sich in Erwerbstätigkeit befinden, erreichen höhere Werte der Lebensqualität als Patienten mit niedrigem Schulabschluss und erwerbslose Patienten. Ein geringerer Leidensdruck, gemessen mit dem FPTM, eine bessere therapeutische Beziehung in der ersten Phase der Behandlung sowie ein höherer Ausgangswert im SF-8 Psych. (entsprechen einer höheren Lebensqualität zu Behandlungsbeginn) gehen mit einer höheren Lebensqualität zur Katamnese einher. Die untersuchten Kliniken und die Prozessmerkmale haben hingegen keinen bedeutenden Vorhersagewert für das sekundäre Outcome. Für die fokussierten explorativen Verlaufsgruppen zeigt sich, dass die Zugehörigkeit zur Non-Response_{EG}-Gruppe eine signifikant geringere katamnestische Lebensqualität vorhersagt [Beta = -.120; p = .014]. Für die Delayed-Response_{EG}-Gruppe kann eine Tendenz der Vorhersage in die gleiche Richtung festgestellt werden [Beta = -.082; p = .084]. Vergleicht man die standardisierten Beta-Koeffizienten, stellt sich der Ausgangswert der Lebensqualität als stärkster Prädiktor der katamnestischen Lebensqualität dar [Beta = .228; p < .001].

Insgesamt erreicht der Zuwachs an erklärter Varianz in der katamnestic Lebensqualität durch die explorativen Verlaufsgruppen lediglich eine kleine Effektstärke [$R^2_{\text{inkrementell}} = 0.02$; $f^2 = 0.030$].

Die in *Hypothese 9* postulierte Annahme, dass durch die *explorativen Verlaufsgruppen* ein signifikanter inkrementeller Anteil der Varianz im sekundären Outcome erklärt wird, kann angenommen werden. Die Nullhypothese wird abgelehnt, da der inkrementelle Varianzanteil sich bedeutend von Null unterscheidet.

Der durch die explorativen Verlaufsgruppen erklärte Varianzanteil erreicht die Höhe einer kleinen Effektstärke.

5.10 Vorhersage des sekundären Outcomes durch die theoriegeleiteten Verlaufsgruppen (Fragestellung 3 D)

5.10.1 Unterscheiden sich die theoriegeleiteten Verlaufsgruppen im sekundären Outcome?

Zur Untersuchung der theoriegeleiteten Verlaufsgruppen hinsichtlich Unterschiede in der katamnestic Lebensqualität wurde eine Kovarianzanalyse mit dem Ausgangswert der Lebensqualität als Kovariate durchgeführt. Die Effektstärken der Veränderung der Lebensqualität von der Aufnahme zur Katamnese für die theoriegeleiteten Verlaufsgruppen und die Ergebnisse der Kovarianzanalyse sind in Tabelle 45 zusammengefasst.

Tabelle 45: Ergebnisse der Aufnahme-Katamnese-Vergleiche der *gesundheitsbezogenen Lebensqualität (SF-8 Psych.)* zwischen den *theoriegeleiteten Verlaufsgruppen* (ANCOVA; $n = 412$)

	Aufnahme (t0)	Katamnese (t2)	Effekt- stärke <i>d</i>	ANCOVA	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	t0 – t2	Faktor: theoriegeleitete Verlaufsgruppe	Kovariate: Ausgangs- belastung
Gesamtstichprobe ($n = 412$)	32.7 (12.1)	40.3 (12.5)	-0.63 ⁺⁺		
Early Response _{TG} ($n = 200$)	28.5 (10.1)	38.4 (11.9)	-0.81 ⁺⁺⁺		
Late Response _{TG} ($n = 39$)	31.7 (12.2)	41.7 (13.7)	-0.82 ⁺⁺⁺	$F(3,407) = 0.594$ $p = .620$; $\eta^2 = .004$	$F(1,407) = 73.302$ $p < .001^{***}$; $\eta^2 = .153^{+++}$
Non-Response _{TG} ($n = 111$)	38.7 (12.7)	43.2 (12.5)	-0.37 ⁺		
n/a ($n = 62$)	36.0 (11.6)	40.5 (12.6)	-0.37 ⁺		

M: Mittelwert; *SD*: Standardabweichung; *p*: Signifikanz; η^2 : Effektstärke Eta²; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**), und große (***) Effektstärken interpretiert.

Anmerkung: Effektstärke *d* berechnet aus $(M_{\text{Prä}} - M_{\text{Kat}}) / SD_{\text{Prä}}$ Gesamtstichprobe

Auch für die theoriegeleiteten Verlaufsgruppen ist in allen Subgruppen ein Anstieg der Lebensqualität von der Aufnahme zur Katamnese zu verzeichnen. Die Veränderung in der Non-Response_{TG}-Gruppe erreicht eine kleine Effektstärke, in der Early-Response_{TG}- und Late-Response_{TG}-Gruppe hingegen jeweils große Effektstärken. Die Ergebnisse der Kovarianzanalyse weisen jedoch nicht auf bedeutsame Unterschiede zwischen den theoriegeleiteten Verlaufsgruppen hin [$F(3,407) = 0.594$; $p = .620$; $\eta^2 = .004$]. Wie aus Abbildung 31 ersichtlich wird, verlaufen die theoriegeleiteten Verlaufsgruppen relativ parallel zueinander. Der Zusammenhang zwischen dem Ausgangswert des SF-8 Psych. und dem katamnestischen SF-8-Psych.-Wert wird durch den signifikanten Effekt der Kovariate Ausgangswert verdeutlicht [$F(1,407) = 73.302$; $p < .001^{***}$; $\eta^2 = .153^{+++}$]. In Abbildung 32 sind zusätzlich die Mittelwerte der Lebensqualität zur Aufnahme, Entlassung und zur Katamnese für die theoriegeleiteten Verlaufsgruppen dargestellt.

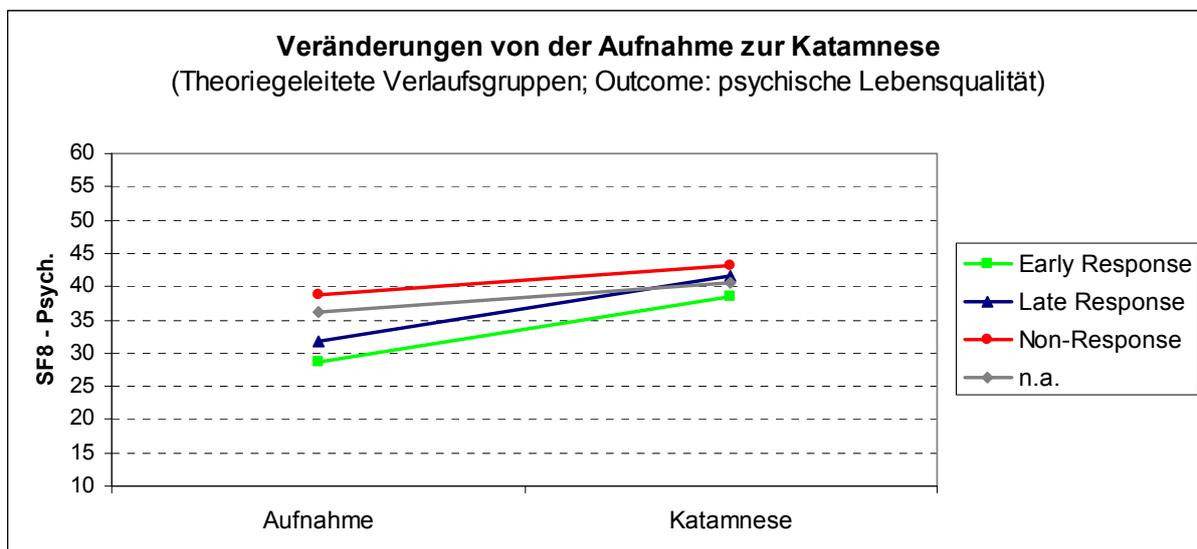


Abbildung 31: SF-8-Psych.-Mittelwerte der theoriegeleiteten Verlaufsgruppen zur Aufnahme und Katamnese

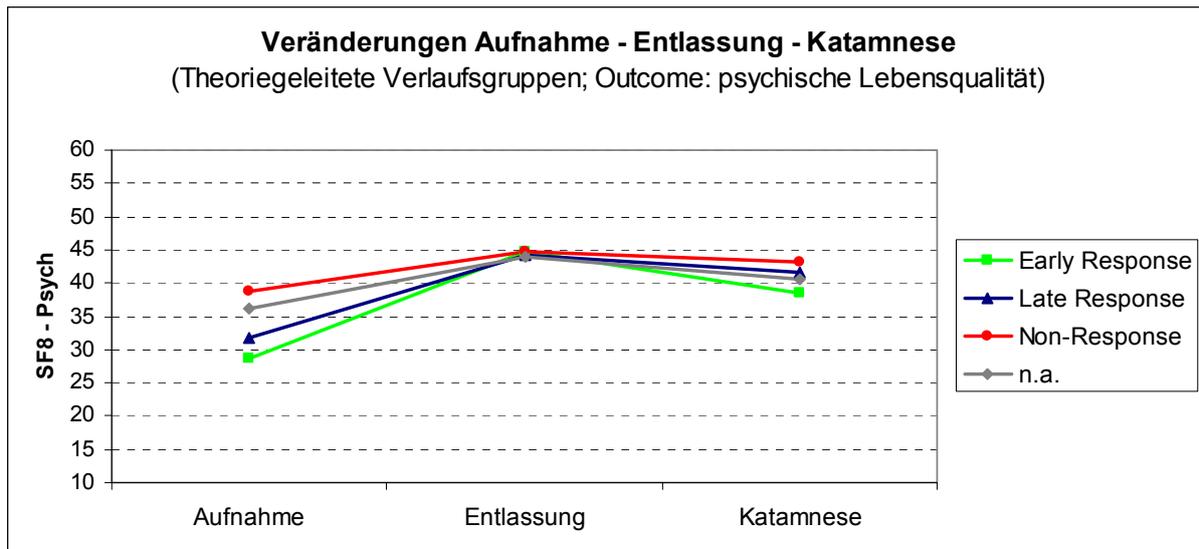


Abbildung 32: SF-8-Psych.-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme, Entlassung und Katamnese

Wie schon die Ergebnisse der Kovarianzanalyse über alle theoriegeleiteten Verlaufgruppen hindeuten erwies sich keiner der paarweisen Vergleiche der katamnestischen Lebensqualität zwischen den Gruppen, unter Konstanthaltung des Ausgangswerts, als statistisch signifikant (vgl. Tabelle 46).

Tabelle 46: Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der gesundheitsbezogenen Lebensqualität (SF-8 Psych.) zwischen jeweils zwei theoriegeleiteten Verlaufgruppen (ANCOVAs)

Theoriegeleitete Verlaufgruppe			Non-Response	Early Response	Late Response
Early Resonse	SF-8 t0	p	< .001***		
		η^2	.134***		
	Gruppe	p	.641		
		η^2	.001		
Late Response	SF-8 t0	p	< .001***	< .001***	
		η^2	.223***	.080**	
	Gruppe	p	.400	.295	
		η^2	.005	.005	
n/a	SF-8 t0	p	< .001***	< .001***	< .001***
		η^2	.267***	.111***	.211***
	Gruppe	p	.451	.632	.167
		η^2	.003	.001	.019 ⁺

p : Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Die in *Hypothese 10* postulierte Annahme, dass sich die theoriegeleiteten Verlaufgruppen, bei statistischer Kontrolle des Ausgangswerts des Ergebnisparameters, bedeutsam im

sekundären Outcome *katamnestisch erfasste gesundheitsbezogene Lebensqualität (psychische Aspekte)* unterscheiden, kann *nicht* angenommen werden. Die Nullhypothese eines nicht signifikanten Unterschieds wird beibehalten.

5.10.2 Auswahl der zu kontrollierenden Prädiktorvariablen

Obwohl die Ergebnisse der Kovarianzanalysen keinen Hinweis auf den Zusammenhang von theoriegeleiteten Verlaufgruppen und sekundärem Outcome liefern, wurde eine hierarchische Regression mit Berücksichtigung von Patienten- und Prozessmerkmalen durchgeführt. Der Vorhersagewert der theoriegeleiteten Verlaufgruppen kann im Zusammenspiel mit anderen Merkmalen einen anderen Stellenwert einnehmen.

In die Regressionsanalyse zur Vorhersage der katamnestischen Lebensqualität durch die theoriegeleiteten Verlaufgruppen wurde das gleiche Prädiktorenset an Patienten- und Prozessmerkmalen in der gleichen Abfolge wie bei der Vorhersage der Lebensqualität durch die explorativen Verlaufgruppen (vgl. Abschnitt 5.9.2) getestet. Die im Fokus stehenden theoriegeleiteten Verlaufgruppen bilden dabei den letzten Block, um den inkrementellen Anteil an erklärter Varianz zu ermitteln.

5.10.3 Vorhersage des sekundären Outcomes durch die theoriegeleiteten Verlaufgruppen

Es liegen keine Hinweise auf Multikollinearität (T : 0.44 – 0.96; VIF : 1.04 – 2.30) oder Autoregression (Durbin-Watson-Koeffizienten = 2.27) in der Regressionsanalyse vor. In Tabelle 47 sind die Ergebnisse zur Modellgüte der hierarchischen Regressionsanalyse zusammengefasst.

Tabelle 47: Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Lebensqualität (SF-8 Psych. zu t2)* durch die *theoriegeleiteten Verlaufgruppen* ($n = 362$)

Modell	Prädiktorenblock	R^2	Änderungsstatistik	
			inkrementelles R^2	p
1	Patientenmerkmale	.30	.30	< .001***
2	1 + Klinik	.30	< .01	.325
3	2 + Prozessmerkmale	.31	< .01	.550
4	3 + theor. Verlaufgruppen	.32	.01	.142

p : Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (***) und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktorenblöcke (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Die Parameter der Modelle 1 bis 3 fallen ähnlich aus wie die der Regressionsanalyse in Abschnitt 5.9.3: Die Patientenmerkmale erklären 30% der Varianz in der katamnestischen Lebensqualität. Klinik- und Prozessvariablen tragen keinen bedeutenden Anteil zur

Varianzaufklärung im Outcome bei. Die hier untersuchten theoriegeleiteten Verlaufsgruppen klären einen Varianzanteil von 1% auf, der Zuwachs erweist sich jedoch als nicht signifikant [$R^2_{\text{inkrementell}} = .011$; $p = .142$]. Der Varianzanteil, der durch das gesamte Modell erklärt wird, beträgt 32%.

Betrachtet man die Parameterschätzer der einzelnen Prädiktoren (s. Tabelle 48), zeigt sich ebenfalls für alle fünf untersuchten Patientenmerkmale ein signifikanter Vorhersagewert: Ein hoher Schulabschluss, Erwerbstätigkeit, ein geringer Leidensdruck zu Behandlungsbeginn (FPTM), eine positive therapeutische Beziehung in der frühen Behandlungsphase und eine höhere Lebensqualität zu Behandlungsbeginn gehen mit einer besseren katamnestischen Lebensqualität einher. Die Klinikvariablen tragen, ebenso wie die Merkmale Behandlungsdauer, Dosis und Art psychotherapienaher Maßnahmen, nicht signifikant zur Vorhersage des sekundären Outcomes bei. Von den theoriegeleiteten Verlaufsgruppen lässt sich für die Gruppe Late Response_{TG} die Tendenz eines Zusammenhangs zum sekundären Outcome feststellen: Die Zugehörigkeit zu dieser Verlaufsgruppe sagt tendenziell eine bessere katamnestische Lebensqualität vorher (Referenzgruppe: Early Response_{TG}) [Beta = .089; $p = .058$]. Den stärksten Prädiktor der katamnestischen Lebensqualität stellt hier die *Leidensdruck*-Skala des FPTM dar [Beta = -.259; $p < .001$].

Tabelle 48: Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der *katamnestischen Lebensqualität (SF-8 Psych. zu t2)* durch die *theoriegeleiteten Verlaufgruppen (n = 362)*

Prädiktor	Beta-Koeffizient	p
Patientenmerkmale		
Schulabschluss hoch	.125	.007**
Erwerbstätigkeit	.172	< .001***
FPTM Leidensdruck	-.259	< .001***
Therapeutische Beziehung	.156	.001***
Ausgangswert SF-8 Psych.	.234	< .001***
Klinik		
Klinik 1	.064	.254
Klinik 3	-.079	.178
Prozessmerkmale		
Behandlungsdauer	.051	.367
Dosis	-.083	.217
Art	.010	.839
Theoriegeleitete Verlaufgruppen		
Late Response _{TG}	.089	.058°
Non-Response _{TG}	.003	.950
n/a	-.051	.313

p: Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**) und dem 0.1%-Niveau (***) festgelegt.

Anmerkung: Die signifikanten Prädiktoren (5%-, 1%- oder 0.1%-Niveau) sind grau unterlegt.

Insgesamt erreicht der Zuwachs an erklärter Varianz in der katamnestischen Lebensqualität durch die theoriegeleiteten Verlaufgruppen nicht das Ausmaß einer kleinen Effektstärke [$R^2_{\text{inkrementell}} = .01$; $f^2 = 0.016$].

Die in *Hypothese 11* postulierte Annahme, dass durch die theoriegeleiteten Verlaufgruppen ein signifikanter inkrementeller Anteil der Varianz im sekundären Outcome erklärt wird, kann *nicht angenommen* werden. Die Nullhypothese wird beibehalten, da der inkrementelle Varianzanteil sich nicht bedeutend von Null unterscheidet.

5.11 Welche Verlaufgruppen erlauben eine bessere Vorhersage des katamnestischen Outcomes?

Zur Überprüfung der letzten Fragestellung der vorliegenden Arbeit wurden die Ergebnisse der Fragestellung 3 für die explorativen Verlaufgruppen und die theoriegeleiteten Verlaufgruppen gegenüberstellend zusammengefasst und ausschließlich deskriptiv verglichen. Die Ergebnisse zum inkrementellen Beitrag der Verlaufgruppen an der Varianzaufklärung im primären und sekundären katamnestischen Outcome sind in Tabelle

49 zusammenfassend dargestellt. Ergänzend ist in der letzten Zeile von Tabelle 49 der Vergleich der Ergebnisse des inkrementellen Beitrags durch die jeweiligen Verlaufsgruppen bei Einschluss der initialen Symptombelastung in das erste getestete Modell dargestellt.

Tabelle 49: Vergleich der *explorativen* und *theoriegeleiteten Verlaufsgruppen* hinsichtlich des Zuwachses an aufgeklärter Varianz im *primären* und *sekundären Outcome* zu t2

Outcome	Explorative Verlaufsgruppen			Theoriegeleitete Verlaufsgruppen		
	$R^2_{\text{inkrem.}}$	p	f^2	$R^2_{\text{inkrem.}}$	p	f^2
Symptombelastung (HEALTH-49 A)	.15	< .001***	.333***	< .01	.336	.014
Lebensqualität (SF-8 Psych.)	.02	.031*	.030 ⁺	.01	.142	.016
Symptombelastung (HEALTH-49 A) ¹	.06	< .001***	.133 ⁺	.02	.015*	.032 ⁺

$R^2_{\text{inkrem.}}$: inkrementelles R^2 ; f^2 : Effektstärke Cohens f^2 ; p : Signifikanz. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (*), mittlere (**), und große (***) Effektstärken interpretiert.

Anmerkung: ¹ Ergebnisse, die unter der Bedingung der Kontrolle der Ausgangsbelastung (Ausgangsbelastung wurde jeweils im ersten Modell berücksichtigt), erzielt wurden.

Es zeigt sich, dass die Varianz im primären Outcome Symptombelastung, auch unter Berücksichtigung weiterer relevanter Prädiktoren, zu einem bedeutenden inkrementellen Anteil von 15% durch die explorativen Verlaufsgruppen erklärt werden kann. Der Zuwachs an erklärter Varianz entspricht einer großen Effektstärke. Der durch die theoriegeleiteten Verlaufsgruppen erklärte Varianzanteil liegt unter 1% und erreicht nicht das Ausmaß einer kleinen Effektstärke.

Die Varianz im sekundären Outcome Lebensqualität wird durch die explorativen Verlaufsgruppen mit 2% zu einem geringen, statistisch aber signifikanten Anteil erklärt, entsprechend einer kleinen Effektstärke. Für die theoriegeleiteten Verlaufsgruppen erreicht der Anteil an Varianzaufklärung im sekundären Outcome mit 1% nicht die statistische Signifikanz.

In der letzten Zeile von Tabelle 49 sind die Ergebnisse der nachgeschalteten Analysen dargestellt, bei denen die Ausgangsbelastung schon in das erste getestete Modell einbezogen wurde (vgl. Abschnitt 5.7.3 und Abschnitt 5.8.3). Selbst wenn die initiale Symptombelastung aus den Verlaufsgruppen auspartialisiert wird, erklären die explorativen Verlaufsgruppen einen bedeutenden inkrementellen Varianzanteil von 6%. Allerdings wird hier der Unterschied zu den theoriegeleiteten Verlaufsgruppen geringer. Bei vorheriger statistischer Berücksichtigung der Ausgangsbelastung tragen die theoriegeleiteten Verlaufsgruppen einen bedeutenden Anteil von 2% zur Varianzaufklärung in der katamnesticen Symptombelastung bei.

Zusammenfassend zeigt die (rein beschreibende) Gegenüberstellung der beiden Gruppen eine Überlegenheit der explorativen Verlaufgruppen in der Vorhersage der katamnestischen Symptombelastung. Auch unter statistischer Kontrolle der Ausgangsbelastung ist dieser Zusammenhang noch aufzufinden. Für die gesundheitsbezogene Lebensqualität fallen die Unterschiede weniger stark aus, der Zuwachs an erklärter Varianz erreicht jedoch nur für die explorativen Verlaufgruppen statistische Signifikanz und die Höhe einer kleinen Effektstärke.

6 Diskussion

In der vorliegenden Arbeit wurde den Fragen zu (1) distinkten Symptomverläufen in der stationären psychotherapeutischen Behandlung, (2) Prädiktoren des Symptomverlaufs und in der Hauptfragestellung (3) zur Vorhersage des längerfristigen Behandlungsergebnisses durch den stationären Symptomverlauf nachgegangen.

Die Diskussion der Ergebnisse untergliedert sich in drei Abschnitte: Im ersten Abschnitt wird auf die Generalisierbarkeit der Ergebnisse eingegangen, indem die Repräsentativität und Selektivität der Untersuchungsstichproben und des untersuchten Settings diskutiert werden (6.1). Im zweiten Abschnitt werden die Ergebnisse der drei Fragestellungen diskutiert. Dafür erfolgt jeweils vorab eine kurze Zusammenfassung der Ergebnisse der einzelnen Fragestellungen. Anschließend werden die Ergebnisse unter Einbezug des bisherigen Forschungsstandes diskutiert (6.2). Nach der inhaltlichen Diskussion werden methodische Einschränkungen der Untersuchung und der erzielten Ergebnisse erörtert (6.3). Im letzten Abschnitt des Kapitels (6.4) wird ein Ausblick in Bezug auf weiterführende Forschungsfragestellungen sowie klinische und versorgungsbezogene Implikationen dargestellt.

6.1 Generalisierbarkeit der Ergebnisse

6.1.1 Studienteilnahme

In einem Erhebungszeitraum von 23 Monaten wurden in den drei untersuchten Kliniken insgesamt 2 399 Patienten behandelt. Nach Prüfung der Ausschlusskriterien standen potentiell 1 847 Patienten für die Untersuchung zur Verfügung. Ausgeschlossen wurden Patienten, die das Erwachsenenalter noch nicht erreicht hatten. Dieses Kriterium diente einer Erhöhung der Repräsentativität hinsichtlich der Behandlung erwachsener Patienten. Das zweite Ausschlusskriterium war eine für dieses Setting sehr kurze oder sehr lange Behandlungsdauer. Das Kriterium der Behandlungsdauer lag bei 3 bis 9 Wochen. Diese Grenzen wurden gewählt, da der Großteil der Behandlungen im untersuchten Setting in dieser Zeitspanne liegt. Aus den Analysen eines Jahresdatensatzes aller über die Deutsche Rentenversicherung finanzierten stationären Behandlungen psychischer Erkrankungen folgt, dass die reguläre Behandlungsdauer bei 96% der Patienten zwischen 3 und 9 Wochen liegt, mehr als die Hälfte aller Patienten werden zwischen 4 und 6 Wochen behandelt (eigene Analysen; Datensatz vgl. Barghaan et al., 2005; s. auch 2.1.2). Mit dem Ausschluss der Patienten mit längeren Behandlungszeiten sollte vermieden werden, Patienten bzw. Indikationsgruppen, die weniger repräsentativ für das untersuchte Setting sind,

auszuschließen. Darüber hinaus hätte der Einschluss von sehr langen und sehr kurzen Behandlungen in den GMM-Analysen Verzerrungen hinsichtlich der identifizierten Gruppen zur Folge haben können. Da die Arbeit darauf abzielt, Aussagen über die reguläre stationäre Behandlung abzuleiten, wurde über dieses Ausschlusskriterium versucht, Behandlungsabbrüche und vorzeitige Entlassungen auszuschließen. Aufgrund der Schwierigkeit der Definition und Erfassung von Behandlungsabbrüchen wurde in der vorliegenden Arbeit auf das von einigen Autoren vorgeschlagene Abbruchkriterium einer begrenzten Behandlungsdauer zurückgegriffen (vgl. Abschnitt 4.2.3). Die Grenze von 3 Wochen ist an das Vorgehen von Damke und Koechel (2002) angelehnt, die den Behandlungsabbruch über alle Entlassungen vor Ablauf von 20 Behandlungstagen definieren. Insgesamt wurden 15% der Klinikpopulation aufgrund einer zu kurzen oder zu langen Behandlungsdauer ausgeschlossen. Dabei wiesen 11% der Klinikpopulation eine Behandlungsdauer von weniger als 3 Wochen und 4% eine Dauer von mehr als 9 Wochen auf. Vorzeitige Entlassungen sind in der Rehabilitation psychischer/psychosomatischer Erkrankungen bei etwa 8% bis 15% der Patienten vorzufinden (Lang, 2003; Lang, Schulz, Lotz-Rambaldi & Koch, 1999), je nach Breite der Definition des Behandlungsabbruchs bzw. der vorzeitigen Entlassung. In der stationären Akutbehandlung werden Abbruchquoten von 20% bis 30% berichtet (Lieberz & Ciemer, 2000). Die hier angenommene Abbrecherquote liegt damit in einer zu erwartenden Höhe. Dennoch muss berücksichtigt werden, dass in den vorliegenden Daten nicht sichergestellt war, dass es sich bei den Behandlungen, die weniger als 3 Wochen andauerten, auch tatsächlich um Behandlungsabbrüche handelte. Aus datenschutzrechtlichen Gründen war es nicht möglich, weitergehende Informationen über die Basisdokumentation der Kliniken hinaus zu erhalten, sodass (potentielle) Behandlungsabbrüche lediglich über die untere Dauergrenze erfasst werden konnten. Sollte ein Teil der Behandlungen regulär so kurz angesetzt worden sein, wäre von einer Selektion hinsichtlich der Behandlungsprozesse auszugehen. Für die Untersuchungsstichproben konnte sichergestellt werden, dass es sich bei allen Patienten um reguläre Entlassungen handelte. In folgenden Studien wäre es wünschenswert, diese Absicherung auch für die ausgeschlossenen Patienten vorzunehmen. Die Ergebnisse der vorliegenden Arbeit sind ausschließlich auf Behandlungen mit einer Dauer von 3 bis 9 Wochen und regulären Entlassungen übertragbar. Trotz dieser Einschränkung sind die Untersuchungsstichproben hinsichtlich der Behandlungsdauer als weitgehend repräsentativ für das Setting (Rehabilitation psychischer/psychosomatischer Erkrankungen und psychosomatisch/psychotherapeutische Krankenhausbehandlung) zu werten, da dieses Spektrum der Behandlungsdauer für den größten Teil aller Behandlungen im untersuchten Setting vorzufinden ist (vgl. Abschnitt 2.1.2).

Von 1 847 Patienten, die potentiell für eine Studienteilnahme zur Verfügung standen, konnten 640 Patienten für die Teilnahme gewonnen werden. Einen Grund für die geringe Teilnahmequote von 35% liegt möglicherweise in dem hohen Aufwand, mit dem die wöchentlichen Erhebungen verbunden waren. Denkbar wäre, dass aus diesem Grund stärker belastete Patienten nicht an der Studie teilgenommen haben. Dies könnte eine systematische Selektion der erreichten Stichprobe hinsichtlich der initialen Belastung zur Folge haben. Die Untersuchungsstichprobe wäre entsprechend weniger stark belastet als die Population, aus der sie stammt. Um mögliche Repräsentativitätseinbußen zu überprüfen, wurden die teilnehmenden Patienten im Hinblick auf bestimmte Merkmale mit allen Patienten, die im Erhebungszeitraum in den drei Kliniken behandelt wurden, verglichen. Hier zeigten sich hinsichtlich des Alters, der Geschlechterverteilung und der Behandlungsdauer keine bedeutenden Unterschiede. Allerdings bestanden Unterschiede in den vergebenen Erstdiagnosen (Entlassdiagnosen): In der Untersuchungsstichprobe finden sich mehr Patienten mit der Erstdiagnose einer Anpassungsstörung und weniger Patienten mit der Erstdiagnose einer unipolaren Depression als in der Klinikpopulation. Anpassungsstörungen mit vorwiegend depressiver Stimmung weisen klassifikatorisch eine hohe Überschneidung mit depressiven Störungen auf und werden dementsprechend auch als Subkategorie der depressiven Störungen diskutiert (Baumeister, 2008). Da die Anpassungsstörung nicht die Kriterien einer anderen spezifischen Störung erfüllen darf, kann diese von der unipolaren Depression unter anderem durch die geringere Schwere und Dauer der Erkrankung abgegrenzt werden (vgl. Kriterien der ICD-10; Dilling et al., 2000). Dieses Ergebnis kann demnach als weiterer Hinweis für die Annahme eines Selektionseffektes hinsichtlich der initialen Krankheits- und Symptomschwere verstanden werden. Letztlich können jedoch nur Vermutungen aufgestellt werden, da von den nichtteilnehmenden Patienten keine weiteren Informationen (z.B. initiale Belastung) vorlagen. Für die Untersuchungsstichproben war es allerdings möglich, die initiale Symptombelastung mit einer Referenzgruppe von $N = 1\,523$ Patienten aus der psychosomatischen Rehabilitationsbehandlung zu vergleichen (Rabung et al., 2008). Die Symptombelastung in der Referenzgruppe wurde ebenfalls bei der Aufnahme in der Klinik erhoben. Die Verlaufsstichprobe unterschied sich nicht bedeutend von der Referenzgruppe in Hinblick auf die initiale Symptombelastung. Lediglich die Katamnese-Stichprobe wies erwartungsgemäß eine signifikant niedrigere initiale Symptombelastung auf. Der Unterschied erreichte jedoch auch hier nicht die Höhe einer kleinen Effektstärke (vgl. Abschnitt 5.1.3). Demzufolge kann angenommen werden, dass die Untersuchungsstichproben hinsichtlich der initialen Symptombelastung repräsentativ für die Population der Patienten der stationären psychosomatischen Rehabilitationsbehandlung ist. Unklar bleibt dennoch, welche weiteren Faktoren zu einer Entscheidung für oder gegen die Studienteilnahme geführt haben. In Abschnitt 5.1.3 wurde dargestellt, dass deutliche

Unterschiede zwischen den drei Kliniken in Hinblick auf die Studienabbrecherquoten bestanden. Lässt sich die Weiterführung der Studienteilnahme anhand von Klinikmerkmalen (bspw. Kommunikation der Studie in der Klinik, Motivation des beteiligten Personals etc.) erklären, könnte die geringere Teilnahmequote auch auf diese externen Parameter zurückgeführt werden. Denkbar wäre auch die Interaktion zwischen Klinik- und Patientenmerkmalen.

6.1.2 Drop-Out

Von den 640 Patienten, die zunächst einer Studienteilnahme zugestimmt hatten, verblieben noch 576 Patienten mit einer ausreichenden Anzahl an Verlaufsdaten in der Untersuchungsstichprobe. Mit 10% fällt der Drop-Out im Verlauf sehr gering aus. In den Drop-Out-Analysen kommt es in keinem der untersuchten Merkmale zu Selektionseffekten. Ein halbes Jahr nach der Entlassung aus der Klinik konnten von 424 Patienten Katamnese-Daten erhoben werden. Der Drop-Out zur Katamnese liegt mit 25% in einem für das untersuchte Setting niedrigen Bereich. Die Katamnese-Stichprobe wurde ebenfalls hinsichtlich möglicher Selektionseffekte untersucht. Es zeigte sich, dass die Patienten der Katamnese-Stichprobe (Verlaufs-Outcome-Stichprobe) initial etwas geringer belastet waren als die Patienten, die zum Katamnese-Zeitpunkt nicht mehr teilgenommen haben (sowohl hinsichtlich der Symptombelastung als auch der gesundheitsbezogenen Lebensqualität). Dies ist einerseits in Bezug auf die Repräsentativität von Bedeutung, die Ergebnisse zum längerfristigen Outcome müssen unter der Annahme interpretiert werden, dass es sich im Hinblick auf das initiale Belastungsniveau um eine selektive Stichprobe handelt. Zum Anderen ist bei der Interpretation zu berücksichtigen, dass die Ergebnisse der Fragestellungen 1 und 2 auf einer höher belasteten Patientenstichprobe basieren als die Ergebnisse der dritten Fragestellung. Zur Beantwortung der ersten Fragestellung erfolgte eine Einteilung der Verlaufsstichprobe in mehrere Gruppen, die unter anderem anhand der Ausgangsbelastung gebildet wurden. Für die dritte Fragestellung wurden die Gruppen mit dem längerfristigen Outcome in Bezug gesetzt. Da hochbelasteten Patienten zur Katamnese gar nicht mehr teilgenommen haben, ist von einer Verzerrung der Ergebnisse der Verlaufs-Outcome-Analysen auszugehen. Dieser Umstand ist auch im höheren Drop-Out in den drei hochbelasteten explorativen Verlaufgruppen (Early Response_{EG}, Delayed Response_{EG} und Non-Response_{EG}) wiederzufinden, während die Gruppe mit einer Linear Response_{EG}, in der die mittlere initiale Symptombelastung am geringsten ausfällt, zur Katamnese stärker repräsentiert ist. Letztlich kann keine Aussage zum längerfristigen Outcome der „verlorenen“ Patienten gemacht werden. Die Verschiebung in Richtung einer geringer belasteten Outcome-Stichprobe könnte aber auch als konservative Schätzung ausgelegt werden: Den gefundenen Unterschieden im Outcome würde demnach, selbst bei einer geringeren

Differenzierung im oberen Belastungsspektrum, eine höhere Bedeutung zukommen. Da sich die Verlaufs-Outcome-Stichprobe nur in wenigen anderen Merkmalen von der Katamnese-Drop-Out-Stichprobe unterscheidet, kann die Stichprobe als weitgehend repräsentativ für die Verlaufsstichprobe gesehen werden.

Zu beachten ist, dass die anfänglich hohe Symptombelastung nicht zwingend mit der Schwere oder Chronifizierung der Erkrankung einhergeht. In den krankheitsbezogenen Merkmalen, die auf eine Chronifizierung hinweisen (Arbeitsunfähigkeitszeiten, psychotherapeutische Vorbehandlungen), bestehen keine Unterschiede zwischen der Verlaufs-Outcome- und der Katamnese-Drop-Out-Stichprobe. Schlussfolgernd muss dieser Selektionseffekt bei der Interpretation der Ergebnisse zum Outcome berücksichtigt werden, es ist jedoch davon auszugehen, dass es eher zu einer Unterschätzung als zu einer Überschätzung des Zusammenhangs zwischen den Verlaufsgruppen und dem Behandlungsergebnis kommt.

In der theoriegeleiteten Gruppeneinteilung ist keine Selektion hinsichtlich des Drop-Outs zur Katamnese zu verzeichnen. Berücksichtigt man den schwächeren Einfluss der Ausgangsbelastung bei der Gruppenbildung (im Vergleich zu der explorativen Gruppenbildung) sowie die über fast alle anderen Merkmale nicht vorhandenen Unterschiede zwischen Verlaufs-Outcome- und Katamnese-Drop-Out-Stichprobe, ist dieses Ergebnis wenig überraschend. Für die Ergebnisse der theoriegeleiteten Verlaufsgruppen kann ebenfalls eher von einer Unter- als Überschätzung des Zusammenhangs der Verläufe mit dem längerfristigen Ergebnis ausgegangen werden.

Zusammenfassend kann die Repräsentativität für beide Untersuchungsstichproben mit wenigen Einschränkungen als gegeben bewertet werden. Die erreichte Teilnahmestichprobe betrug zwar lediglich 35% aller in Frage kommenden Patienten, diese unterschied sich aber nur in der Verteilung der Patienten mit der Erstdiagnose einer depressiven bzw. Anpassungsstörungen von der Klinikpopulation. In der Geschlechterverteilung, dem Alter, der Behandlungsdauer und allen anderen Erstdiagnosen lagen keine wesentlichen Unterschiede zwischen Teilnehmern und Nichtteilnehmern vor. Über Selektionseffekte hinsichtlich nicht untersuchter Variablen, wie beispielsweise der initialen Symptombelastung, kann keine Aussage gemacht werden. Von den teilnehmenden Patienten, für die durch die Baseline-Erhebung ein breites Spektrum an Merkmalen vorlag, brachen nur 10% die Studienteilnahme im Verlauf der Behandlung ab. Drop-Out-Analysen erbrachten keine bedeutenden Unterschiede zwischen der Untersuchungsstichprobe und der Drop-Out-Stichprobe. Zur Katamnese konnten noch 74% der Patienten der Untersuchungsstichprobe erreicht werden, wobei die Katamnese-Stichprobe ein höheres Alter und eine geringere

initiale Belastung im Vergleich zur Drop-Out-Stichprobe aufwies. In allen anderen untersuchten Merkmalen fanden sich keine bedeutenden Unterschiede.

Die Untersuchung wurde multizentrisch in drei Fachkliniken durchgeführt, um die Repräsentativität der Ergebnisse zu erhöhen. Zwei der drei Kliniken sind der medizinischen Rehabilitation psychischer/psychosomatischer Erkrankungen zuzuordnen, während die dritte Klinik im Versorgungsbereich der psychosomatischen/psychotherapeutischen Krankenhausbehandlung angesiedelt ist. Folglich kann eine höhere Repräsentativität für die stationäre rehabilitative Behandlung als für die Krankenhausbehandlung angenommen werden. Die in der somatischen Medizin üblicherweise verwendeten Kriterien zur Abgrenzung zwischen Rehabilitations- und Krankenhausbehandlung sind allerdings nur sehr eingeschränkt auf den Bereich der Behandlung von psychisch erkrankten Patienten übertragbar (Schulz et al., 2008). Einen Schwerpunkt in der Rehabilitationsbehandlung bilden die Maßnahmen zur Reintegration psychisch erkrankter Patienten. Da aber viele psychische Erkrankungen mit Einschränkungen der psychosozialen Funktionsfähigkeit einhergehen, stellen die Maßnahmen zur Wiederherstellung psychosozialer Funktionen auch einen wesentlichen und unverzichtbaren Anteil der Krankenhausbehandlung dar (Schulz & Koch, 2002). Die Abgrenzung der beiden Versorgungsbereiche sowohl hinsichtlich der Behandlungsmodalitäten und -prozesse als auch im Hinblick auf die behandelten Diagnosegruppen ist nur schwer vorzunehmen. Allerdings lag in zwei der drei untersuchten Kliniken (Krankenhaus- und Rehabilitationsbehandlung) ein Schwerpunkt auf der Behandlung von Essstörungen. Der Anteil der Patienten mit einer Essstörung als Erstdiagnose lag in den Untersuchungsstichproben (über alle drei Kliniken) bei 7% und 6%, in der Klinikpopulation bei 8%. Diese Häufigkeiten liegen zwar deutlich über dem Anteil von ca. 2% an Patienten mit der Erstdiagnose einer Essstörung in der Population der Rehabilitation psychischer/psychosomatischer Störungen (Deutsche Rentenversicherung Bund, 2005), entspricht aber weitgehend der Häufigkeit dieser Diagnosegruppe in der psychosomatischen/psychotherapeutischen Krankenhausbehandlung (ca. 6%; vgl. Schulz et al., 2008). Demzufolge kann hier von der Übertragbarkeit der Ergebnisse auch auf die psychosomatische/psychotherapeutische Krankenhausbehandlung ausgegangen werden. Einschränkend bleibt die Eingrenzung der Behandlungsdauer und die potentiell geringere initiale Symptombelastung der untersuchten Patienten im Vergleich zur Population, auf die die Ergebnisse übertragen werden sollen, zu berücksichtigen.

6.2 Diskussion der Ergebnisse der Fragestellungen und Hypothesen

6.2.1 Welche Behandlungsverläufe der Symptombelastung lassen sich identifizieren?

In der *ersten Fragestellung* wurde der Verlauf der Symptombelastung während der stationären Behandlung untersucht und geprüft, ob sich homogene Symptomverlaufgruppen identifizieren lassen. Die *explorative Vorgehensweise (Fragestellung 1 A)* resultierte in fünf distinkten Verlaufgruppen, die sich anhand der Ausgangsbelastung und der Veränderung über den gesamten Behandlungsverlauf unterscheiden. Die Gruppen wurden anhand der Verlaufsparemeter inhaltlich beschrieben und als Linear Response_{EG}, Early Response_{EG}, Delayed Response_{EG}, Non-Response_{EG} und Atypical Course benannt. Die Linear-Response_{EG}-Gruppe stellt mit 71% der Patienten die größte Subgruppe dar. Die verbleibenden vier Verlaufgruppen weisen Häufigkeitsanteile von 4% bis 11% auf. In dem *theoriegeleiteten Vorgehen (Fragestellung 1 B)* wurden vier Verlaufgruppen anhand des Zeitpunktes und des Ausmaßes der Symptomverbesserung vorab definiert. Die Hälfte der untersuchten Patienten wies nach den theoriegeleiteten Kriterien eine Early Response_{TG} auf. In die Gruppe der Late Response_{TG} fallen 10% der Patienten, und weitere 24% der Patienten erfüllen die Kriterien für beide Gruppen nicht und bilden die Non-Response_{TG}-Gruppe. Die verbleibenden 16% können aufgrund fehlender Daten nicht zugeordnet werden. Im Vergleich der beiden Verlaufgruppenbildungen zeigen sich erhebliche Unterschiede, die sich nicht nur in der Passung der äquivalenten Gruppen, sondern schon in der Anzahl und Größe der gefundenen Gruppen abbilden. Die in *Hypothese 1* formulierte Annahme, dass mindestens 5% der Patienten den Verlauf einer theoriegeleiteten Early Response aufweisen, konnte angenommen werden.

Die Ergebnisse der explorativen Verlaufsanalyse korrespondieren mit den bisherigen Befunden zur Verlaufsanalyse der Symptomatik psychisch erkrankter Patienten. Wie in den in Abschnitt 2.5.7 dargestellten Studien konnten mehrere distinkte Subgruppen gefunden werden (Cinar et al., 2010; Keller & Hautzinger, 2007; Lutz et al., 2007a; Lutz et al., 2009; Stulz et al., 2007; Uher et al., 2010). Die in den bisherigen Verlaufsuntersuchungen häufige gefundene Gruppe mit einer frühen Verbesserung konnte auch in der vorliegenden Studie identifiziert werden. Dabei zeichnen sich die Patienten mit diesem Verlauf durch eine mittlere bis hohe Ausgangsbelastung aus, was mit den Ergebnissen aus den Untersuchungen der Arbeitsgruppe von Lutz et al. korrespondiert (Lutz et al., 2009; Lutz et al., 2007b; Stulz et al., 2007). In den meisten Studien zur Identifikation von Subgruppen des Symptomverlaufs lässt sich zumindest eine Verlaufgruppe identifizieren, die eine frühe Verbesserung im Behandlungsverlauf abbildet (Cinar et al., 2010; Keller & Hautzinger, 2007; Lutz et al., 2009; Stulz et al., 2007; Uher et al., 2010).

Ähnlich wie in der Studie von Keller und Hautzinger (2007) wurde in der vorliegenden Arbeit neben mehreren relativ kleinen Subgruppen eine größere Patientensubgruppe mit einem kontinuierlichen Verbesserungsverlauf gefunden. Die größte Gruppe der von Keller und Hautzinger untersuchten depressiven Patienten (insgesamt $n = 203$) zeichnete sich durch eine mittlere Ausprägung der depressiven Symptomatik zu Behandlungsbeginn und eine langsame aber stetige Verbesserung aus. Diese Gruppe ist mit einem Anteil von 65% und dem beschriebenen Verlaufsmuster vergleichbar mit der hier gefundenen Linear-Response_{EG}-Gruppe. Die zweitgrößte Gruppe (26%) wird, ähnlich wie die hier identifizierte Non-Response-Gruppe, mit einer hohen Ausgangsbelastung und einer geringen Veränderung der depressiven Symptombelastung beschrieben. Die verbleibenden beiden Verlaufsgruppen, die sich in der depressiven Stichprobe identifizieren ließen, können ebenfalls äquivalent zu den hier gefundenen Gruppen Early Response (3% bei Keller und Hautzinger) und Delayed Response (6% bei Keller und Hautzinger) beschrieben werden. Trotz der unterschiedlichen Häufigkeitsverteilungen innerhalb der identifizierten Gruppen ist eine gewisse Übereinstimmung in der Anzahl der gefundenen Gruppen sowie hinsichtlich der Art der Verläufe zu verzeichnen. Diese Ergebnisse scheinen insbesondere im Hinblick auf das untersuchte Setting von Bedeutung zu sein: In der Studie von Keller und Hautzinger werden, anders als in den meisten Studien zum Symptomverlauf, auch stationäre Verläufe eingeschlossen (Keller & Hautzinger, 2007). Allerdings werden keine differenzierten Angaben über die Anteile der Patienten aus den unterschiedlichen Behandlungssettings oder zu einem möglichen Zusammenhang zwischen Setting und Verlauf gemacht. Letztlich können die Ergebnisse der explorativen Verlaufsanalyse nur schwer eingeordnet werden, da die empirische Grundlage spezifisch für das hier untersuchte stationäre Setting bislang fehlt. Eine kritische Auseinandersetzung mit dem statistischen Verfahren des Growth Mixture Modeling, das zur Identifikation der Verlaufsgruppen angewendet wurde, erfolgt gesondert in Abschnitt 6.3.3.

Die Ergebnisse der explorativen Verlaufsgruppenbildung liefern erste wertvolle Hinweise über potentielle Symptomverlaufsgruppen in der stationären psychotherapeutischen Behandlung (psychosomatisch/psychotherapeutische Rehabilitations- und Krankenhausbehandlung, vgl. Abschnitt 6.1). Eine differenzierte Betrachtung der einzelnen Verlaufsgruppen erfolgt in der Diskussion der Ergebnisse der Prädiktoren- und Outcome-Analysen in den folgenden Abschnitten.

Auch für die theoriegeleitete Methode der Verlaufsgruppenbildung lassen sich Übereinstimmungen mit einigen Befunden finden. Der Anteil der Patienten mit einer Early Response_{TG} fällt in der vorliegenden Untersuchung mit 50% hoch aus. Zwar finden auch andere Autoren nach ähnlichen Kriterien relativ hohe Anteile zwischen 37% und 43% an früh

verbesserten Patienten (Kelly et al., 2005; Tang & DeRubeis, 1999b; Tang et al., 2002), in anderen Studien werden jedoch auch geringere Quoten (19%) an Early-Response-Verläufen gefunden (Stiles et al., 2003). Die Anzahl der Patienten, die sich früh im Verlauf verbessern, variiert stark in Abhängigkeit von der angewandten Definition einer frühen Verbesserung. In den beschriebenen Studien variieren die Kriterien sowohl hinsichtlich des Zeitpunkts als auch des Mindestausmaßes der zu erreichenden Verbesserung. Ferner unterscheiden sich die untersuchten Stichprobengrößen in den Studien stark, was die Vergleichbarkeit der Ergebnisse erschwert.

Für die Gruppen Late Response_{TG} und Non-Response_{TG} ließen sich in der vorliegenden Untersuchung ebenfalls substantielle Anteile finden. Hinsichtlich der Häufigkeitsanteile dieser beiden theoriegeleiteten Verlaufgruppen konnten keine Hypothesen formuliert werden, da die theoretische Fundierung hierzu fehlt (vgl. Abschnitt 2.5). Empirische Vergleichswerte können nur schwer herangezogen werden, weil die Definition und Operationalisierung der Verlaufgruppen sehr uneinheitlich verwendet werden (vgl. Abschnitt 2.5.4). Die unterschiedlichen methodischen Vorgehensweisen müssen sowohl bei den Ergebnissen der theoriegeleiteten als auch der explorativen Gruppenbildung berücksichtigt werden. Auf diesen Aspekt wird im nächsten Abschnitt, in dem die explorative und die theoriegeleitete Gruppenbildung vergleichend diskutiert werden, näher eingegangen.

Vergleich der explorativen und theoriegeleiteten Verlaufgruppen

Die explorativen Verlaufgruppen wurden nach deutlich unterschiedlichen Kriterien (Einbezug der Ausgangsbelastung und des gesamten Verlaufs) gegenüber den theoriegeleiteten Verlaufgruppen (partielle Unabhängigkeit von der Ausgangsbelastung³⁴, zeitlich punktuelle Kriterien der (Non-)Response) gebildet (vgl. Abschnitt 4.3.4 und 4.5.2). Diese Unterschiede in den Kriterien finden sich auch in den resultierenden Gruppen wieder. Trotz der hier gewählten identischen bzw. ähnlichen Terminologie weisen die jeweils (scheinbar) äquivalenten Gruppen eine geringe Übereinstimmung auf. Alleine die erzielte Gruppenanzahl und die jeweiligen Gruppengrößen zeigen schon deutlich die fehlende Übereinstimmung. Lediglich die explorative Early-Response_{EG}-Gruppe findet sich zu 94% in der theoriegeleiteten Early-Response_{TG}-Gruppe wieder. Auffällig ist hier auch, dass die theoriegeleiteten Early-Response_{TG}-Patienten in den explorativen Verlaufgruppen zu einem substantiellen Anteil der Non-Response_{EG}-Gruppe zugeordnet wurden.

³⁴ Für die theoriegeleitete Verlaufgruppenbildung wurde die Ausgangsbelastung zwar einbezogen, indem die Differenzwerte zwischen der Belastung zu Beginn der Behandlung und der Belastung zu einem kritischen Zeitpunkt als Effektstärken der Veränderung berechnet wurden. Die absolute Höhe der Symptombelastung stellte jedoch kein Kriterium zur Einteilung der theoriegeleiteten Verlaufgruppenbildung dar.

Gleichzeitig fiel der größte Teil aller theoriegeleiteten Verlaufsgruppen in die explorative Verlaufsgruppe Linear Response_{EG}, die jedoch keine Entsprechung in dieser Einteilung hat. Die sich stark unterscheidenden Häufigkeiten in den einzelnen Verlaufsgruppen zwischen den beiden Einteilungen erklären dieses Ergebnis: Während mit der explorativen Vorgehensweise durch den Einschluss des gesamten Verlaufs eher eine Ausdifferenzierung stark abweichender Verläufe von einer großen Gruppe mit einem „Standardverlauf“ stattfand, resultierte die Gruppenbildung nach theoretisch hergeleiteten Kriterien, die ausschließlich punktuelle Verbesserungen einbezog, in einer Verteilung, in der die Early Response_{TG} die größte Gruppe darstellt. Allerdings befindet sich in dieser Gruppe auch die Hälfte der Patienten, die in den explorativen Verlaufsgruppen als Linear Response_{EG} zugeteilt wurden. Das bedeutet, dass die Hälfte der Linear-Response_{EG}-Patienten das Kriterium einer theoriegeleiteten Early Response_{TG} erfüllt, und sogar ein weiteres Viertel der Linear-Response_{EG}-Gruppe der theoriegeleiteten Non-Response_{TG} zugeordnet wurde. Hier wird ein Vorteil der explorativen Vorgehensweise deutlich: Die Patienten der Linear-Response_{EG}-Gruppe, die in die theoriegeleiteten Non-Response_{TG}-Gruppe fallen ($n = 109$), werden bei der theoriegeleiteten Vorgehensweise nicht als (Linear) Responder „entdeckt“. Dies ist möglicherweise darauf zurückzuführen, dass eine stetige Verbesserung der Symptombelastung stattfand, das Kriterium aber nach der zweiten oder dritten Woche (noch) nicht erreicht wurde. Zur Entlassung könnte ein punktueller Anstieg der Symptombelastung erneut das Erreichen des Kriteriums verhindert haben. Die explorative Vorgehensweise ist daher unempfindlicher gegenüber Schwankungen bzw. Ausreißern, solange diese nicht zu stark ausgeprägt sind. Andererseits kann bei der explorativen Vorgehensweise nicht für jeden Patienten von einer bestimmten (Mindest-)Verbesserung zu einem oder mehreren Zeitpunkten im Verlauf ausgegangen werden. Die hohe Anzahl an nicht zuzuordnenden Patienten in der theoriegeleiteten Definition der Verlaufsgruppen macht einen weiteren Vorteil der explorativen Verlaufsanalyse mittels GMM deutlich. Mit dem Verfahren werden alle vorhandenen Messzeitpunkte ausgeschöpft, sodass fehlende Werte bei einzelnen Messzeitpunkten nicht so stark ins Gewicht fallen und nicht unbedingt ganze Fälle aus der Zuordnung auszuschließen sind.

Betrachtet man das von Tang und DeRubeis beschriebene Phänomen der *Sudden Gains* (Tang & DeRubeis, 1999b) wird deutlich, dass plötzlich auftretende, starke Veränderungen der Symptombelastung mit der theoriegeleiteten Vorgehensweise nicht mit Sicherheit „entdeckt“ werden können. In der explorativen Vorgehensweise würde ein solches Phänomen mit höherer Wahrscheinlichkeit (im Sinne einer Zuordnung zu einer anderen als der Standardgruppe) entdeckt werden. Für weiterführende Forschungsansätze wäre auch eine Kombination aus den beiden Vorgehensweisen vorstellbar. In einem ersten Schritt könnte man die Gruppen explorativ identifizieren und anschließend anhand vorab definierter

Kriterien überprüfen. Hierbei liefe man jedoch auch Gefahr, dass es zu einer zu großen Aufspaltung der Stichprobe kommt und Aussagen über prototypische Verläufe erschwert werden.

Ist nun die mittlere Verlaufskurve über mehrere Messzeitpunkte im Behandlungsverlauf bedeutender für die Bestimmung von Symptomverläufen oder das Erreichen eines bestimmten Zielkriteriums zu einem festgelegten Zeitpunkt? Hier schließt sich direkt die Frage an, mit welchem Ziel und zu welchem Zweck eine Einteilung der Symptomverläufe vorgenommen wird. Ziel der vorliegenden Studie war eine möglichst hohe Vorhersagbarkeit des längerfristigen Behandlungsergebnisses (Fragestellung 3), an der auch die Güte der Gruppeneinteilung beurteilt werden soll. Die Bedeutung der Bildung von Verlaufgruppen wird entsprechend in den folgenden Abschnitten im Zusammenhang mit der Vorhersagbarkeit des Outcomes diskutiert (Abschnitt 6.2.3). Anschließend werden die beiden Gruppenbildungen unter Berücksichtigung der Ergebnisse vergleichend bewertet. Die vertiefende Diskussion methodischer Aspekte der Verlaufgruppenbildung erfolgt gesondert in Abschnitt 6.3.3, auf die Angemessenheit der wöchentlichen Erhebungen zur Verlaufgruppenbildung wird in Abschnitt 6.3.5 eingegangen.

6.2.2 Welche Merkmale erlauben eine Vorhersage des Behandlungsverlaufs?

In der *zweiten Fragestellung* wurden die *Hypothesen 2* und *3* untersucht, dass ein bedeutender Anteil der Varianz jeweils in den explorativen und theoriegeleiteten Verlaufgruppen durch theoretisch fundierte Patientenmerkmale erklärt wird. Zur *Vorhersage der explorativen Verlaufgruppen* wurden die Wahrscheinlichkeiten der Gruppenzugehörigkeit zur Early-Response_{EG}-, Delayed-Response_{EG}- und Non-Response_{EG}-Gruppe berechnet. Die Referenzgruppe bildete die Linear-Response_{EG}-Gruppe. Zusammenfassend zeigen die Ergebnisse folgendes Bild: Der Verlauf einer Early Response_{EG} kann mit einer hohen Zuverlässigkeit durch einen erhöhten Leidensdruck zu Behandlungsbeginn vorhergesagt werden. Tendenziell haben Patienten mit einer somatoformen Störung und Patienten mit psychischer Komorbidität eine erhöhte Chance einer Early Response_{EG}. Ferner gehen niedrige Werte der Selbstwirksamkeitserwartung und hohe Werte der Verneinung psychischer Hilfsbedürftigkeit mit einer erhöhten Wahrscheinlichkeit einer Early Response_{EG} einher. Außer für den Leidensdruck sind die Effekte der Prädiktoren jedoch nur gering ausgeprägt und als Tendenz zu bewerten. Deutlicher sind die Effekte der Prädiktoren der Zugehörigkeit zur Delayed-Response_{EG}-Gruppe. Ein hoher Leidensdruck und die (Ko-)Diagnose einer Persönlichkeitsstörung gehen jeweils mit einer stark erhöhten Wahrscheinlichkeit einer Delayed Response_{EG} einher. Längere Arbeitsunfähigkeitszeiten erhöhen auch bedeutend die Wahrscheinlichkeit dieses Verlaufs. Tendenzielle Effekte zeigen sich hier für die Prädiktoren niedriger Schulabschluss

und höheres Alter. Für die Gruppe der Non-Response_{EG} lassen sich fünf bedeutende Prädiktoren finden: Patienten mit einem niedrigen Schulabschluss, längeren Arbeitsunfähigkeitszeiten, höherem Leidensdruck, psychischer Komorbidität, einer bestehenden Partnerschaft und nicht erwerbstätige Patienten haben jeweils eine erhöhte Wahrscheinlichkeit, im Laufe der Behandlung *nicht* zu respondieren. Patienten mit einer ambulanten Vorbehandlung weisen hingegen eine geringere Wahrscheinlichkeit einer Non-Response_{EG} auf.

Hinsichtlich der *Vorhersage der theoriegeleiteten Verlaufgruppen* wurden die Wahrscheinlichkeiten der Zugehörigkeit zu den beiden Gruppen Late Response_{TG} und Non-Response_{TG} untersucht, die Early-Response_{TG}-Gruppe stellte die Referenzgruppe dar. Zusammenfassend lässt sich die Gruppenzugehörigkeit zu einem Verlauf mit einer Late Response_{TG} am stärksten durch längere Arbeitsunfähigkeitszeiten vor Behandlungsbeginn, das Vorhandensein einer diagnostizierten Angststörung und durch eine geringere Ausprägung der Ausgangsbelastung vorhersagen. Weniger deutlich fallen die Ergebnisse zur Komorbidität aus: Diese geht tendenziell mit einer geringeren Wahrscheinlichkeit einer Late Response_{TG} einher. Für die (Ko-)Diagnosen Persönlichkeitsstörung und Adipositas zeigt sich eine Tendenz für den Verlauf einer Late Response_{TG}. Ebenso erweist sich der Vorhersagewert einer geringeren Selbstwirksamkeitserwartung für die Zugehörigkeit zur Late-Response_{TG}-Gruppe lediglich als tendenziell. Die Zugehörigkeit zur Non-Response_{TG}-Gruppe wird durch längere Arbeitsunfähigkeitszeiten, die (Ko-)Diagnose einer Persönlichkeitsstörung, aber auch durch psychotherapeutische ambulante Vorbehandlungen vorhergesagt. Längere Behandlungszeiten gehen ebenfalls verstärkt mit einem Non-Response_{TG}-Verlauf einher. Hingegen tragen eine positive Einschätzung der Güte der therapeutischen Beziehung und eine hohe Ausgangsbelastung zu einer verringerten Wahrscheinlichkeit einer Non-Response_{TG} bei. Tendenzuelle Effekte für die Eintrittswahrscheinlichkeit einer Non-Response_{TG} weisen eine bestehende Partnerschaft sowie die (Ko-)Diagnose einer somatoforme Störung auf.

Insgesamt erwies sich der Anteil der durch die Patientenmerkmale erklärten Varianz in beiden Verlaufgruppen als signifikant, *Hypothese 2* und *Hypothese 3* konnten angenommen werden. Die Höhe des erklärten Varianzanteils erreichte mit 50% und 51% jeweils eine große Effektstärke.

6.2.2.1 *Vorhersage des Symptomverlaufs anhand von Patientenmerkmalen*

Im Folgenden werden die Ergebnisse der Prädiktoranalysen im Einzelnen anhand der relevanten Prädiktoren diskutiert. Dabei wird jeweils sowohl auf die explorativen als auch auf die theoriegeleiteten Verlaufgruppen Bezug genommen. Die folgenden Abschnitte behandeln die Prädiktoren *initiale Symptombelastung* und *psychischer Leidensdruck*,

Diagnose und Komorbidität, Psychotherapiemotivation, Partnerschaft, Bildungsniveau und Erwerbsstatus sowie die Güte der therapeutischen Beziehung.

Symptombelastung, Leidensdruck und Diagnose als Prädiktoren des Symptomverlaufs

Die initiale Symptombelastung wurde nur zur Vorhersage der theoriegeleiteten Verlaufgruppen, nicht aber für die Vorhersage der explorativen Verlaufgruppen einbezogen. Die Begründung für dieses Vorgehen wurde bereits in Abschnitt 5.6.3.2 erläutert. In den theoriegeleiteten Verlaufgruppen war die *Symptombelastung* ein relevanter Prädiktor für eine erhöhte Wahrscheinlichkeit einer Non-Response_{TG} und einer Late Response_{TG}. Vergleichbar zu den Ergebnissen der theoriegeleiteten Verlaufgruppen zeigte sich in den explorativen Verlaufgruppen die Skala *psychischer Leidensdruck* als zum Teil hoch signifikanter Prädiktor für die Gruppen Early Response_{EG}, Delayed Response_{EG} und Non-Response_{EG}. Die Skala *Leidensdruck* ist eine Skala des Fragebogens zur Psychotherapiemotivation (FPTM-23), die über die folgenden vier Items gebildet wird: „*Ich leide stark unter seelischen Problemen*“, „*Ich fühle mich seelisch angegriffen und belastet*“, „*Ich werde mit meinen Problemen nicht mehr alleine fertig*“ und „*Ich habe Angst vor der Zukunft*“. Es wird deutlich, dass die Inhalte der Items zwar allgemeiner formuliert sind, jedoch nur schwer zur Symptombelastung abgegrenzt werden können. Dies zeigt auch die Untersuchung der Testgütekriterien des FPTM-23, in der die Korrelationen der Skala *Leidensdruck* mit den Skalen der SCL-90-R (Franke, 2002) mit $r = .58$ für die Ängstlichkeitsskala und $r = .71$ für die Depressivitätsskala hoch ausfallen (Schulz et al., 2003). Damit übereinstimmend wurde auch in den vorliegenden Daten eine relativ hohe Korrelation der Skala *Leidensdruck* mit der über den HEALTH-49 A gemessenen *Symptombelastung* gefunden ($r = .60$; s. Anhang: Tabelle A.4). Folglich können die Ergebnisse zum *psychischen Leidensdruck* weitgehend äquivalent zur *initialen Symptombelastung* interpretiert werden.

In Anbetracht der hohen Korrelation der beiden Skalen und der Grundlage, auf der die explorativen Verlaufgruppen gebildet wurden (also unter Einbezug der initialen Symptombelastung), ist der hohe Prädiktorwert des psychischen Leidensdrucks für die explorativen Verlaufgruppen wenig überraschend. Patienten der Gruppen Early Response_{EG}, Delayed Response_{EG} und Non-Response_{EG} weisen eine höhere initiale Symptombelastung auf als die Linear-Response_{EG}-Gruppe. Folglich geht ein hoher psychischer Leidensdruck wahrscheinlicher mit dem Verlauf der drei hoch belasteten Gruppen einher. In den theoriegeleiteten Verlaufgruppen erwies sich eine *niedrige* initiale Symptombelastung als bedeutender Prädiktor für die Gruppen Late Response_{TG} und Non-Response_{TG}, verglichen mit der Wahrscheinlichkeit einer Early Response_{TG}. Hier steht demnach eine hohe Ausgangsbelastung im Zusammenhang mit einer frühen Verbesserung

im Verlauf. Für die Interpretation dieser sich scheinbar widersprechenden Ergebnisse sollen auch die univariaten Vergleiche ebenso wie weitere Prädiktoren der Verlaufsgruppen herangezogen werden.

Betrachtet man die univariaten Vergleiche zwischen den beiden Gruppenbildungen, so fällt auf, dass die Ergebnisse wenig Übereinstimmung aufweisen. Während in den explorativen Gruppen vor allem die Linear-Response_{EG}-Gruppe mit einer geringen Symptombelastung, einem höheren Bildungsniveau und weniger chronifizierten Erkrankungen einhergeht, treten die Unterschiede in den theoriegeleiteten Verlaufsgruppen in eine andere Richtung auf: Hier ist vor allem die Non-Response_{TG}-Gruppe durch eine geringe Symptombelastung zu Beginn der Behandlung, durch einen höheren Anteil an erwerbstätigen Patienten und geringere Chronifizierung der Erkrankungen gekennzeichnet. Dahingegen scheinen die Patienten der theoriegeleiteten Early-Response_{TG}-Gruppe, die zu Beginn der Behandlung die höchste Symptombelastung aufweisen, vorerst ungünstigere Voraussetzungen für ein Ansprechen auf die Behandlung mitzubringen. Wie kann die dennoch auftretende frühe Verbesserung bei diesen Patienten (Early Response_{TG}), im Vergleich zum ausbleibenden Ansprechen bei den Patienten der Non-Response_{TG}-Gruppe erklärt werden? Ein Erklärungsansatz liegt in einer methodischen Einschränkung der theoriegeleiteten Gruppenbildung: Um zu einer Response-Gruppe zugeordnet zu werden, müssen die Patienten ein bestimmtes Ausmaß an Verbesserung der Symptombelastung erreichen, unabhängig vom absoluten Wert der Ausgangsbelastung. Es zeigt sich jedoch, dass in der Verlaufsstichprobe von 576 Patienten knapp ein Fünftel der Patienten ($n = 112$; 19%) bei Aufnahme in der Klinik unterhalb des Cut-Off-Wertes der entsprechenden Skala liegen. Das bedeutet, dass diese Patienten, zumindest zum Zeitpunkt der Baseline-Messung, nur sehr gering belastet sind. Der Baseline-Wert wird als Referenz für eine relevante Veränderung herangezogen. Dies schließt einerseits die Frage nach der Sensitivität des Instruments bzw. der Skala für den unteren Wertebereich ein, andererseits können Patienten, die zum ersten Messzeitpunkt einen sehr niedrigen Skalenwert aufweisen sich rein rechnerisch nicht um die hier definierte Größe verbessern. Dies trifft auf 36 Patienten (Skalenwert zu t_0 unterhalb von 0.317), also 5% der Verlaufsstichprobe, zu. Diese 36 Patienten können den theoriegeleiteten Verlaufsgruppen Early Response_{TG} und Late Response_{TG} selbst bei einer Verbesserung zur vollständigen Remission (ausbleibende Symptomatik mit dem Skalenwert 0) nicht zugeordnet werden. Selbst wenn eine Verbesserung im definierten Ausmaß rechnerisch möglich wäre, scheint es auch inhaltlich fraglich, ob davon auszugehen ist, dass sich Patienten mit einer geringen Belastung über einen kurzen Zeitraum im gleichen Ausmaß verbessern können bzw. müssen (um einer Response-Gruppe zugeteilt zu werden), wie Patienten mit einer höheren Belastung.

Einen inhaltlichen Erklärungsansatz liefern die in den multivariaten Analysen identifizierten weiteren Prädiktoren der theoriegeleiteten Verlaufgruppen: Eine Non-Response_{TG} wird durch längere Behandlungszeiten, längere Arbeitsunfähigkeitszeiten, eine geringere Güte der therapeutischen Beziehung, eine komorbide Persönlichkeitsstörung sowie ambulante psychotherapeutische Vorbehandlung vorhergesagt. Lange Arbeitsunfähigkeitszeiten, eine geringe Ausgangsbelastung und die geringe Güte der therapeutischen Beziehung stellen die stärksten Prädiktoren der theoriegeleiteten Non-Response_{TG} dar. Die Merkmale längere Arbeitsunfähigkeitszeiten, komorbide Persönlichkeitsstörung sowie ambulante psychotherapeutische Vorbehandlung können als Indikatoren für eine Chronifizierung der Erkrankung gesehen werden (Steffanowski et al., 2007). In einer der wenigen Studien, in denen Prädiktoren von Symptomverlaufgruppen untersucht wurden, zeigt sich übereinstimmend mit den vorliegenden Ergebnissen die Kodiagnose einer Persönlichkeitsstörung als relevanter Prädiktor des Verlaufs einer Non-Response (Cinar et al., 2010). Die geringe Güte der therapeutischen Beziehung könnte im Zusammenhang mit der Komorbidität einer Persönlichkeitsstörung stehen. Es wäre denkbar, dass sich der Beziehungsaufbau bei Patienten mit einer komorbiden Persönlichkeitsstörung schwieriger gestaltet. Die übergeordneten diagnostischen Kriterien einer spezifischen Persönlichkeitsstörung umfassen laut ICD-10 die Störung der sozialen Funktionsfähigkeit und deutliche Abweichungen in Beziehungen zu anderen (Dilling et al., 2000). Welche Rolle der therapeutischen Beziehung dabei zukommt kann hier nicht geklärt werden. Vorstellbar wäre ein moderierender Effekt, oder eine Interaktion von der (Ko-)Diagnose einer Persönlichkeitsstörung und der Güte der therapeutischen Beziehung, die eine Non-Response_{TG} begünstigen. Eine kausale Verursachung im Sinne eines Mediatoreffektes kann für die untersuchte Stichprobe ausgeschlossen werden, da die (Ko-)Diagnose einer Persönlichkeitsstörung auch bei gleichzeitiger Untersuchung der therapeutischen Beziehung noch ein relevanter Prädiktor der Non-Response war. Die Dauer der Krankschreibungen wurde schon in vielen Studien als Prädiktor eines schlechteren Behandlungsergebnisses gefunden (Steffanowski et al., 2007; vgl. auch Abschnitt 2.4.3) und wird häufig als Indikator für Chronifizierung verstanden. Die Vorerfahrung mit ambulanter Psychotherapie ist ebenfalls ein Hinweis auf eine längere Erkrankungsdauer bzw. Chronifizierung der Erkrankung. Dabei widerspricht die geringe Ausgangsbelastung dieser Annahme nicht unbedingt, da die akute Symptombelastung gerade bei chronifizierten psychischen Erkrankungen nicht immer im Vordergrund steht (beispielsweise bei der Diagnose einer Dysthymie). Ebenso zeichnen sich Persönlichkeitsstörungen per Definition durch eine andauernde und tiefgreifende Psychopathologie, entsprechend einer Chronifizierung, aus, die jedoch nicht zwingend durch eine akute hohe Symptombelastung gekennzeichnet sein muss.

Zusammenfassend zeigen die Ergebnisse zur Vorhersage der explorativen und theoriegeleiteten Verlaufgruppen, dass ungünstige Verläufe (Non-Response, Delayed Response und Late Response) hinsichtlich krankheitsbezogener Merkmale durch längeren Arbeitsunfähigkeitszeiten, ambulante Vorbehandlung und psychische Komorbidität insbesondere mit einer Persönlichkeitsstörung wahrscheinlicher werden. Diese Merkmale konnten auch in Übersichtsarbeiten als relevante Prädiktoren der Behandlungsdauer gefunden werden (Melchior et al., 2010; Steffanowski et al., 2007). Vergleicht man die Behandlungszeiten zwischen den hier identifizierten Symptomverlaufgruppen, findet man signifikante Gruppenunterschiede (s. Anhang: Tabelle C.3 und Tabelle D.3). In den explorativen Verlaufgruppen werden Patienten mit einer Non-Response_{EG} und einer Delayed Response_{EG} länger behandelt als Patienten mit einer Linear Response_{EG} und Early Response_{EG}. Allerdings weist vor allem die Gruppe *Atypical Course*, die eine Restkategorie darstellt und nicht inhaltlich interpretiert wird, die kürzesten Behandlungszeiten auf. Ähnlich ist es bei den theoriegeleiteten Verlaufgruppen: Zwar werden die Gruppen Non-Response_{TG} und Late Response_{TG} länger behandelt als die Patienten der Early-Response_{TG}-Gruppe, die größte Abweichung zur mittleren Behandlungsdauer ist hier ebenfalls in der Restkategorie *n/a* zu finden. Dies weist schon darauf hin, dass die Unterschiede nicht überinterpretiert werden sollten, was sich auch in den multivariaten Analysen bestätigt. Beim Einschluss weiterer Merkmale trägt die Behandlungsdauer nicht zur Vorhersage der (inhaltlich interpretierbaren) Verlaufgruppen bei. Daraus kann geschlussfolgert werden, dass chronifizierte Patienten eher einen ungünstigen Symptomverlauf aufweisen, der dann zu längeren Behandlungen führt. In Abschnitt 6.2.2.2 wird vertiefend auf die Vorhersage des Symptomverlaufs über die Prozessmerkmale Behandlungsdosis, -art und -dauer eingegangen.

Psychotherapiemotivation als Prädiktor des Symptomverlaufs

In das Vorhersagemodell der Symptomverläufe wurden vier der sechs Skalen des Fragebogens zur Psychotherapiemotivation eingeschlossen. Die verbleibenden zwei Skalen *Wissen* und *Initiative* wurden nicht in das Modell aufgenommen, weil sich die Verlaufgruppen in diesen Skalen univariat nicht bedeutend unterschieden. Der über den Fragebogen zur Psychotherapiemotivation erhobene *psychische Leidensdruck* zeigt sich in den explorativen Verlaufgruppen als bedeutender Prädiktor. Infolge der hohen Korrelation zwischen der Skala *Leidensdruck* mit der Symptombelastung wurden die Ergebnisse dieser Skala im Zusammenhang mit der initialen Symptomschwere diskutiert.

Betrachtet man die Ergebnisse zu den verbleibenden drei Skalen *Verneinung psychischer Hilfsbedürftigkeit*, *Hoffnung* und *symptombezogene Zuwendung durch andere* ergibt sich ein uneinheitliches Bild. Die Verneinung und die symptombezogene Zuwendung konnten als

tendenziell signifikante Prädiktoren für den Verlauf einer Non-Response_{EG} identifiziert werden, allerdings in entgegengesetzte Richtungen. Eine höhere Verneinung der Hilfsbedürftigkeit, die als Anzeichen für geringe Motivation zur Psychotherapie gewertet wird, erhöht ebenso die Chance einer Early Response wie die geringe Ausprägung der symptombezogenen Zuwendung durch andere, die für eine hohe Psychotherapiemotivation steht. Zur symptombezogenen Zuwendung ließen sich allerdings auch in der Validierungsstudie von Schulz et al. (2003) ebenso wie in früheren Auswertungen (Dohrenbusch & Scholz, 2002; Schulz et al., 1995) keine bedeutenden Zusammenhänge mit korrespondierenden Merkmalen finden (z.B. die Fremdeinschätzung der Psychotherapiemotivation durch den Therapeuten oder die selbsteingeschätzte Intention zur aktiven Mitarbeit an der Rehabilitationsmaßnahme). Entsprechend sollen die Ergebnisse dieser Skala vorsichtig interpretiert werden.

Insgesamt lassen sich die inkonsistenten Ergebnisse zur Psychotherapiemotivation wie folgt interpretieren: Ausschließlich der *psychische Leidensdruck* - im Sinne einer hohen Psychotherapiemotivation - erlaubt eine bedeutende Vorhersage für den Verlauf einer Early Response_{EG}, Delayed Response_{EG} und Non-Response_{EG} der explorativen Verlaufgruppen. Dies ist durch den starken Zusammenhang zur initialen Symptombelastung bedingt. Eine stärker ausgeprägte *Verneinung psychischer Hilfsbedürftigkeit*, also eine geringe Psychotherapiemotivation, führt tendenziell wahrscheinlicher zum Verlauf einer Early Response_{EG} oder einer Non-Response_{EG}, während die geringe *Zuwendung durch Andere*, die eine hohe Psychotherapiemotivation anzeigen, wiederum eher zur Non-Response_{EG} führt. Zusammengefasst lassen die Ergebnisse eher auf den Prädiktorwert einer *geringen* Psychotherapiemotivation für eine Early Response_{EG} ebenso wie für eine Non-Response_{EG} schließen. Da sich die beiden Gruppen im Vergleich zur Linear-Response_{EG}-Gruppe durch eine ähnlich hohe Ausgangsbelastung, jedoch sehr unterschiedliche Verläufe auszeichnen, könnte auch ein Interaktionseffekt zwischen Psychotherapiemotivation und initialer Symptombelastung angenommen werden. Demnach würde die Psychotherapiemotivation nicht direkt, sondern vermittelt über eine hohe Ausgangsbelastung mit dem Verlauf der Symptombelastung zusammenhängen. Allerdings handelt es sich lediglich um tendenziell signifikante Prädiktoren, die zudem zumindest für die Early Response_{EG} gegen die erwartete Richtung des Zusammenhangs sprechen. Hinsichtlich der Vorhersage der theoriegeleiteten Verlaufgruppen erwiesen sich die Skalen der Psychotherapiemotivation als nicht bedeutend.

Zum Zusammenhang von Psychotherapiemotivation und Symptomverlauf können aus den Ergebnissen dieser Arbeit keine eindeutigen Schlussfolgerungen gezogen werden. Dieser Aspekt sollte in Folgeuntersuchungen verstärkt fokussiert werden.

Partnerschaft, Bildungsniveau und Erwerbsstatus als Prädiktoren des Symptomverlaufs

Ein eher überraschendes Ergebnis hinsichtlich der Vorhersage des Symptomverlaufs durch Patientenmerkmale stellt die *Partnersituation* dar. In beiden Gruppenbildungen weist eine bestehende Partnerschaft auf eine erhöhte Wahrscheinlichkeit einer Non-Response hin. Grundsätzlich müssen die schon beschriebenen Unterschiede zwischen der explorativen und theoriegeleiteten Non-Response-Gruppen berücksichtigt werden. Dennoch ist dieser Befund von Bedeutung, da dieser Prädiktor zwar eine theoretische Fundierung aufweist, allerdings bestand die Annahme eines umgekehrten Zusammenhangs. Aus den Studien zur ambulanten Psychotherapie geht weitgehend eindeutig hervor, dass eine Partnerschaft mit einem positiven Outcome einhergeht (Clarkin & Levy, 2004). Im stationären Setting lassen sich bislang keine Zusammenhänge zwischen der Partnersituation und dem Behandlungsergebnis finden (Hölzel et al., 2010; Steffanowski et al., 2007). Eine Erklärung für die hier erzielten Ergebnisse eines Zusammenhangs zwischen der Partnerschaft und der Non-Response ist von methodischer Natur: Betrachtet man die Partnersituation in den univariaten Vergleichen der Verlaufsgruppen, lassen sich keine bedeutenden Unterschiede feststellen. Lediglich bei den theoriegeleiteten Verlaufsgruppen sind in der Non-Response_{TG}-Gruppe anteilig die meisten Patienten, die in einer Partnerschaft leben, die Unterschiede zu den anderen Gruppen sind aber marginal. Für die explorativen Verlaufsgruppen zeigte sich sogar, dass in den Gruppen Linear Response_{EG} und Early Response_{EG} mehr Patienten in einer Partnerschaft leben als in der Non-Response_{EG}-Gruppe. Hier wird der Wert der zusätzlichen Prüfung auf multivariate Unterschiede deutlich: Eine bestehende Partnerschaft scheint demnach nicht *per se* mit der Non-Response zusammenzuhängen, sondern vielmehr in Wechselwirkung mit anderen Prädiktoren. Für die explorativen Verlaufsgruppen wäre eine Wechselwirkung mit anderen bedeutenden soziodemographischen Merkmalen wie Schulabschluss oder Erwerbstätigkeit denkbar. Eine Vermutung besteht darin, dass Patienten mit höherem Bildungsniveau ihre Partnerschaft besser als Ressource nutzen können, weil hier die emotionale Befindlichkeit möglicherweise vermehrt verbalisiert wird und Probleme mit dem Partner besprochen werden (Bassler et al., 1995b). Patienten, die diese Möglichkeit weniger oder gar nicht in der Partnerschaft oder dem sozialen Umfeld haben, könnten potentiell auch von der therapeutischen Arbeitsbeziehung weniger profitieren und weniger auf die Behandlung ansprechen (Hahlweg, 2005; Schütz & Wiesner, 2000). Auf der anderen Seite können die Ergebnisse auch dahingehend interpretiert werden, dass die potentielle Schutzfunktion einer Partnerschaft mit der geringeren Symptombelastung zu Behandlungsbeginn zusammenhängt, die wiederum den oben erläuterten Zusammenhang zur Non-Response bedingt (Schiepek & Cremers, 2003). Die Ergebnisse müssen jedoch mit Vorsicht interpretiert werden, da relevante Mediatorvariablen wie etwa die Qualität der Partnerschaft nicht erhoben wurden. Ferner zeigen die Outcome-Analysen, dass eine

bestehende Partnerschaft, unter Berücksichtigung anderer Prädiktoren, in keinem Zusammenhang zum längerfristigen Behandlungsergebnis steht. Die Partnerschaft scheint kein vorrangig bedeutendes Merkmal zur Vorhersage von Verlauf und Outcome zu sein, vielmehr muss die Partnersituation im Zusammenhang mit weiteren Variablen untersucht werden. Schließlich sollten Ergebnisse, die inkonsistent mit bisherigen Befunden sind, aufgrund des explorativen Charakters der Verlaufsanalysen mit einer gewissen Vorsicht interpretiert werden.

Patienten mit einem *niedrigen Bildungsniveau* und/oder *nicht erwerbstätige Patienten* sprechen nur verzögert oder gar nicht auf die Behandlung an. Dies könnte durch die Art der Behandlungsmaßnahmen, die möglicherweise auf ein höheres Bildungsniveau ausgerichtet sind, erklärt werden. Hier wäre auch eine fehlende Passung von therapeutischen Konzepten und Bedarfen dieser Patientengruppe denkbar. Auch könnte das Bildungsniveau der behandelnden Therapeuten und Ärzte eine zu große Diskrepanz zu dem der Patienten aufweisen, sodass Inhalte und Prozesse der Psychotherapie nicht optimal bei diesen Patienten greifen können. Möglicherweise unterscheiden sich auch die Bedürfnisse und Erwartungen dieser Patienten an die Therapeuten von den Vorstellungen der Patienten mit höherem Bildungsniveau (Bassler et al., 1995b). Die Ergebnisse hinsichtlich des Prädiktors Erwerbstätigkeit gehen einher mit dem Befund, dass eine höhere Ausprägung von Problemen im Arbeitsumfeld mit einer geringeren Wahrscheinlichkeit eines günstigen Symptomverlaufs (Early Response) einhergehen (Lutz et al., 2007b). Eine weitere Erklärung für den schlechteren Verlauf der erwerbslosen Patienten könnte in dem spezifischen Bedarf an (mehr) sozialtherapeutischen und berufsbezogenen Inhalten liegen. In den hier durchgeführten Analysen zur Dosis von Behandlungsmaßnahmen sind nur psychotherapienahe Maßnahmen differenziert betrachtet worden, Maßnahmen zur beruflichen Wiedereingliederung wurden nicht analysiert. Es stellt sich die Frage, ob arbeitslose Patienten oder Patienten, die nicht in einem Erwerbsstatus stehen, eine spezifische Problematik mit sich bringen, auf die in den Kliniken nicht ausreichend eingegangen wird.

Die Ergebnisse zum Bildungsniveau und zur Erwerbstätigkeit sind insofern besonders relevant, da epidemiologische Studien eine erhöhte Prävalenz psychischer Erkrankungen bei Menschen mit einem niedrigen sozioökonomischen Status aufzeigen, die gleichzeitig mit einem geringeren Behandlungserfolg einhergeht (Andrade, Caraveo-Anduaga, Berglund, Bijl, Kessler, Demler et al., 2000; Falconnier, 2009; Talala, Huurre, Aro, Martelin & Prattala, 2009). Ein interessanter Befund findet sich auch hinsichtlich des Zusammenhangs der Güte mit der patientenseitigen Zufriedenheit mit der therapeutischen Beziehung: Bassler et al. (1995b) fanden in ihrer Untersuchung zur Testgüte des HAQ, dass Patienten mit niedrigem Schulabschluss eine höhere Zufriedenheit mit der Beziehung zum Therapeuten aufweisen

als Patienten mit hohem Schulabschluss. Da die Daten in der stationären Behandlung erhoben wurden, kann von einer weitgehenden Vergleichbarkeit zur vorliegenden Untersuchung ausgegangen werden. Wie sind die Befunde im Zusammenhang mit den hier gefundenen Ergebnissen zu interpretieren? Die höhere Zufriedenheit mit der Beziehungsgüte könnte über die oben erwähnten unterschiedlichen Erwartungshaltungen hinsichtlich Psychotherapie bei unterschiedlichem Bildungsniveau erklärt werden. Unter der Annahme, dass im sozialen Umfeld der Patienten mit niedrigem Bildungsniveau weniger über Probleme und psychischen Leidensdruck geredet wird, könnte die Beziehung zum Therapeuten im Vergleich zur mitgebrachten Beziehungserfahrung als besonders positiv bewertet werden (Bassler et al., 1995b). Gleichzeitig scheint jedoch alleine die Beziehungszufriedenheit nicht ausreichend zu sein, um eine therapeutische Arbeitsbeziehung aufzubauen, die zu einer kurzfristigen Symptombesserung führt.

Ein weiterer Erklärungsansatz für den Zusammenhang von Bildungsniveau und Non-Response wurde von Hölzel et al. (2010) vorgeschlagen. In den Ergebnissen einer multizentrischen Analyse hinsichtlich der Prädiktoren von Non-Response erwies sich der Schulabschluss nur im univariaten Vergleich zwischen Respondern und Non-Respondern als unterschiedlich. Response wurde hier über eine Verbesserung im BDI (Hautzinger et al., 1994) um mindestens 50% von der Aufnahme zur Entlassung operationalisiert. Unter Einbezug des Prozessmerkmals *erhaltene Psychotherapie* zeigte sich, dass dieser Effekt nicht mehr auftrat. Das schlechtere Behandlungsergebnis der Patienten mit niedrigem Schulabschluss scheint nicht auf das Bildungsniveau, sondern auf einen schlechteren Zugang zu psychotherapeutischen Behandlungsmaßnahmen zurück zu gehen. Da das Outcome in der Studie von Hölzel et al. (2010) zur Entlassung erhoben wurde sind die Ergebnisse vergleichbar mit der hier definierten Gruppe der theoriegeleiteten Non-Response. Übertragen auf die Ergebnisse der Verlaufsanalyse würde das bedeuten, dass Patienten mit niedrigem Bildungsniveau und/oder Erwerbsstatus weniger Psychotherapie erhalten, was zu einem ungünstigen Verlauf der Symptombelastung führt. Allerdings konnte die Annahme des mediierenden Effektes der Dosis psychotherapienaher Maßnahmen nicht überprüft werden, da diese nicht in die Verlaufsanalysen einbezogen wurde. Auf die Rolle der Dosis der psychotherapeutischen Behandlung für den Zusammenhang zwischen sozioökonomischem Status und längerfristigem Behandlungsergebnis wird in der Diskussion der Outcome-Analysen eingegangen.

Anders als in zwei Studien zur Identifikation und Vorhersage von Symptomverlaufgruppen, in denen das Alter einen relevanten Prädiktor des Symptomverlaufs darstellte (Stulz et al., 2007; Uher et al., 2010), konnte in der vorliegenden Untersuchung weder univariat noch multivariat ein Zusammenhang zwischen dem Alter und dem Symptomverlauf gefunden

werden. Entsprechend sollte der Zusammenhang zum Alter in zukünftigen Studien mit untersucht werden.

Güte der therapeutischen Beziehung als Prädiktor des Symptomverlaufs

Die Güte der therapeutischen Beziehung wird in der Literatur als einer der robustesten Prädiktoren der Wirksamkeit von Psychotherapie beschrieben (Beutler et al., 2006; Beutler et al., 2004; Martin et al., 2000; Orlinsky et al., 2004; vgl. auch Abschnitt 2.4.2). In den vorliegenden Analysen unterscheidet sich die patientenseitig eingeschätzte Güte der therapeutischen Beziehung zwischen den explorativen Verlaufgruppen bedeutend. Patienten mit einer Linear Response_{EG} und Patienten mit einer Early Response_{EG} schätzen die Güte der Beziehung in einer frühen Phase der Behandlung bedeutend höher ein als Patienten mit einer Delayed_{EG} oder einer Non-Response_{EG}. Werden allerdings weitere Merkmale berücksichtigt, trägt dieses Merkmal nicht mehr zur Vorhersage der explorativen Verlaufgruppen bei. Offensichtlich stehen die schon genannten Merkmale wie der initiale Leidensdruck oder Indikatoren für Chronifizierung in einem stärkeren Zusammenhang mit dem Symptomverlauf als die Güte der therapeutischen Beziehung. Die Güte der therapeutischen Beziehung wurde in der zweiten und dritten Behandlungswoche erhoben. Es wird jedoch angenommen, dass die therapeutische Beziehung im Sinne eines Prozessmerkmals Veränderungen und Schwankungen im Behandlungsverlauf aufweist (Martin et al., 2000). Aufgrund dessen bleibt auch die Kausalbeziehung zwischen der Güte der therapeutischen Beziehung und der Symptombelastung unklar: Die Einschätzung der Güte der therapeutischen Beziehung kann ebenso von dem Verlauf der Symptombelastung (bis zur zweiten bzw. dritten Woche) abhängen wie der Symptomverlauf von der Beziehung. Zwar wurde die Güte der therapeutischen Beziehung als Prädiktor der Verläufe untersucht, strenggenommen kann aber nur von einem Zusammenhang gesprochen werden. Zur differenzierten Untersuchung des Zusammenhangs zwischen zwei sich verändernden Merkmalen müssten beide Merkmale kontinuierlich im Verlauf erhoben werden.

Neben der punktuellen Messung ist auch der Erhebungszeitpunkt der Güte der therapeutischen Beziehung kritisch zu betrachten. Die Wahl des Messzeitpunktes orientierte sich an den für den HAQ vorgeschlagenen Erhebungszeitpunkt. Im ambulanten Setting wurde der Fragebogen in der Regel nach der dritten Behandlungswoche eingesetzt (Luborsky et al., 1996). Für das stationäre Setting, in dem eine höhere Behandlungsdichte vorzufinden ist, wurde die zweite Behandlungswoche vorgeschlagen (Bassler et al., 1995b). Dennoch bleibt unklar, ob gerade für das breite Behandlungsspektrum im stationären Setting eine zuverlässige Einschätzung nach der zweiten Woche erwartet werden kann. Auf die Validität der Messung der therapeutischen Beziehung in der stationären Behandlung wird vertiefend in der methodischen Diskussion (Abschnitt 6.3.4.3) eingegangen.

6.2.2.2 Vorhersage des Symptomverlaufs anhand von Prozessmerkmalen

Die Prüfung der Vorhersage der Verlaufszugehörigkeit durch *Prozessmerkmale* wirft die Frage nach der Kausalität auf. Da es sich bei der Behandlungsdauer, -dosis und -art um Merkmale handelt, die sich im Behandlungsprozess verändern können, stellt sich die Interpretation hinsichtlich der Richtung der Vorhersage als problematisch dar. Es ist davon auszugehen, dass nicht nur der Symptomverlauf von den Prozessvariablen abhängen kann, sondern auch, dass die Prozessvariablen während der Behandlung an den Verlauf der Symptombelastung des Patienten angepasst werden.

Aus den hier erzielten Ergebnissen geht nicht hervor, welches Ereignis zeitlich früher auftrat, die Veränderung der Symptombelastung oder beispielsweise die Veränderung der Dosis psychotherapienaher Behandlungsmaßnahmen. Um der Frage der Kausalbeziehung näher zu kommen, müsste die Behandlungsdosis ebenfalls im Verlauf untersucht werden. So könnten die Verläufe der Symptombelastung und die Veränderung der Dosis in Hinblick auf systematische Effekte untersucht werden. Nur unter der Annahme, dass die zu Beginn der Behandlung festgelegte Dauer im weiteren Verlauf nicht angepasst werden kann, und unter der Voraussetzung einer kontinuierlich gleichbleibenden Dosis über die gesamte Behandlungszeit hinweg könnte aus der Gesamtdauer bzw. -dosis eine Vorhersage des Verlaufs erfolgen. Selbst bei der eingeschränkten Varianz der Behandlungsdauer im untersuchten Setting und der relativ konstanten Dosis an psychotherapienahen Maßnahmen (vgl. Abschnitt 2.5.1) ist nicht davon auszugehen, dass die Prozesse im stationären Setting diesen starren Strukturen zugrunde liegen.

Behandlungsdauer als Prädiktor des Symptomverlaufs

Die Analyse des Vorhersagewertes der Behandlungsdauer für die Symptomverlaufgruppen erbrachte lediglich für die Zugehörigkeit zur theoriegeleiteten Non-Response_{TG}-Gruppe eine bedeutend erhöhte Wahrscheinlichkeit. Eine längere Behandlung hängt demnach eher mit einer Non-Response_{TG} zusammen. Der Prädiktorwert der Behandlungsdauer kann hier lediglich im Sinne eines Zusammenhangs mit einer Non-Response_{TG} interpretiert werden und nicht als kausale Beziehung (vgl. vorherigen Abschnitt). Ebenso könnte erst der Verlauf einer Non-Response_{TG} zu einer längeren Behandlung geführt haben: Insofern würde es nicht überraschen, dass diese Patienten, die sich im frühen Verlauf der Behandlung nicht verbessert haben und in der Folgedauer auch keine wesentlichen Verbesserungen aufweisen³⁵, länger behandelt werden als Patienten, die sich schon früh im Verlauf

³⁵ Es ist zwar anzunehmen, dass die Patienten der Non-Response_{TG}-Gruppe, die sich bis zur dritten Behandlungswoche und zur Entlassung nicht wesentlich verbessert haben, auch während der gesamten Behandlung keine wesentliche und anhaltende Verbesserung aufweisen. Genau genommen kann aber keine Aussage zum gesamten Verlauf der Symptombelastung gemacht werden,

verbessern. Dies ist im Sinne des Konzepts der *responsive regulation* gut nachvollziehbar (Stiles et al., 2008a; Stiles et al., 1998): Der Behandlungsprozess, also auch die Entscheidung über die Dauer, kann und soll während der Behandlung und unter Berücksichtigung der Zielerreichung angepasst werden. Es ist nur verständlich, dass gerade die Behandlung der Patienten, die im Laufe der Behandlung nicht ansprechen, verlängert wird. Allerdings können diese Annahmen nicht überprüft werden, da mögliche Anpassungen der Behandlungsdauer nicht erfasst werden konnten. Aussagen zur Kausalbeziehung zwischen Symptomverlauf und Behandlungsdauer sind aufgrund des Untersuchungsdesigns nicht möglich.

Dosis psychotherapienaher Behandlungsmaßnahmen als Prädiktor des Symptomverlaufs

In der vorliegenden Untersuchung lag die Dosis psychotherapienaher Behandlungsmaßnahmen lediglich als Gesamtdosis vor, sodass intraindividuelle Unterschiede im Verlauf nicht berücksichtigt werden konnten. In den univariaten Gruppenvergleichen zeigte sich jedoch, dass sich Patienten mit unterschiedlichen Verläufen bedeutend hinsichtlich der erhaltenen Dosis psychotherapienaher Maßnahmen unterschieden. Der Vergleich zwischen den explorativen Verlaufgruppen zeigte, dass die Patienten der Early-Response_{EG}-Gruppe mit durchschnittlich insgesamt 38.6 Stunden die höchste Dosis psychotherapienaher Maßnahmen erhalten haben (Dauer in Minuten: $M = 2\,313$; $SD = 873$). Gefolgt wird diese von der Gruppe der Delayed Response_{EG} mit durchschnittlich insgesamt 37.9 Stunden psychotherapienaher Maßnahmen (Dauer in Minuten: $M = 2\,275$; $SD = 1\,083$). Die Non-Response_{EG}-Gruppe, die die längsten Behandlungsdauern aufweist, erhält während der gesamten Behandlung hingegen eine Dosis von nur 35.8 Stunden (Dauer in Minuten: $M = 2\,149$; $SD = 975$). Die geringste Dosis psychotherapienaher Maßnahmen ist mit durchschnittlich 33.5 Stunden bei der Linear-Response_{EG}-Gruppe zu verzeichnen (Dauer in Minuten: $M = 2\,008$; $SD = 791$). Geht man davon aus, dass sich die psychotherapeutischen Behandlungsmaßnahmen gleichmäßig auf die Behandlungswochen verteilen, so kann in der Early-Response_{EG}-Gruppe von 6.5 Stunden pro Woche³⁶, hingegen in der Non-Response_{EG}-Gruppe von durchschnittlich 5.6 Stunden pro Woche ausgegangen werden. Für die Patienten mit einer Linear Response_{EG} wären es entsprechend auch 5.6 Stunden pro Woche und bei der Delayed Response_{EG} 6 Stunden pro Woche. Die Ergebnisse liefern erste Hinweise auf einen bestehenden Zusammenhang zwischen Behandlungsdosis und Symptomverlauf. Denkbar wären auch ein

da dieser nicht in die Einteilung der theoriegeleiteten Verlaufgruppen einbezogen wurde.

³⁶ Um die Stundenzahl pro Woche zu ermitteln, wurde die Gesamtstundenanzahl der Verlaufgruppe durch die mittlere Behandlungsdauer (in Tagen) der jeweiligen Gruppe dividiert und anschließend mit 7 multipliziert.

Interaktionseffekt zwischen der Behandlungsdosis und weiteren Merkmalen, wie z.B. der Güte der therapeutischen Beziehung, auf den Verlauf der Symptombelastung. Aus den vorliegenden Ergebnissen lassen sich hierzu allerdings keine Aussagen ableiten: Neben der Schwierigkeit der Interpretation hinsichtlich kausaler Zusammenhänge stellte die hohe Anzahl fehlender Werte in den Dosisvariablen eine zweiten Grund dar, weshalb diese in den multivariaten Analysen zur Vorhersage der Gruppenzugehörigkeit nicht berücksichtigt werden konnten. Eine weitere Einschränkung der Ergebnisse ist durch die fragliche Validität der KTL-Daten bedingt (vgl. Abschnitt 6.3.4.2). Weiterführende Untersuchungen des Zusammenhangs zwischen Psychotherapiedosis und Symptomverlauf sowie zur Frage der Kausalität sind notwendig.

6.2.3 Lässt sich das längerfristige Behandlungsergebnis durch den Symptomverlauf während der Behandlung vorhersagen?

Die *dritte Fragestellung* stellt die Hauptfragestellung dar und prüft die Vorhersage des längerfristigen Behandlungsergebnisses durch die jeweiligen Verlaufsgruppen. Dabei wurden für das primäre und das sekundäre Outcome jeweils Hypothesen hinsichtlich vorhandener Gruppenunterschiede sowie eines bedeutenden Zuwachses an Varianzaufklärung im Outcome formuliert.

Die *explorativen Verlaufsgruppen* unterscheiden sich auch unter statistischer Kontrolle der Ausgangsbelastung bedeutsam in der *katamnestischen Symptombelastung*. Es fällt auf, dass der Effekt der Veränderung in der Early-Response_{EG}-Gruppe am größten ist. Vergleicht man die einzelnen Verlaufsgruppen hinsichtlich des Outcomes miteinander, so erweisen sich alle Paarvergleiche bis auf den Vergleich von Delayed und Non-Response_{EG} als signifikant. Die Untersuchung der Vorhersage des primären Outcomes zeigt, dass ein bedeutender Anteil der Varianz durch die Patientenmerkmale Schulabschluss, Erwerbstätigkeit, Güte der therapeutischen Beziehung und die Diagnose einer Anpassungsstörung erklärt wird. Die Prozessmerkmale Dosis und Art der psychotherapeutischen Behandlungen sowie die zu kontrollierende Klinikzugehörigkeit tragen nicht wesentlich zur Vorhersage des primären Outcomes bei. Lediglich die Behandlungsdauer erweist sich als signifikanter Prädiktor. Durch die Hinzunahme der explorativen Verlaufsgruppen wird die Varianzaufklärung im primären Outcome deutlich erhöht. Letztlich leistet auch die Ausgangsbelastung einen signifikanten inkrementellen Beitrag zur Varianzaufklärung. Insgesamt wird eine niedrige katamnestische Symptombelastung in diesem Modell durch einen hohe Schulbildung, eine bestehende Erwerbstätigkeit, die Diagnose einer Anpassungsstörung, eine positive Einschätzung der Güte der therapeutischen Beziehung, längere Behandlungszeiten und die Zugehörigkeit zur Early-Response_{EG}-Gruppe begünstigt. Patienten der Verlaufsgruppen Delayed Response_{EG}

und Non-Response_{EG} und/oder Patienten mit einer hohen Symptombelastung zu Behandlungsbeginn gehen eher mit einem schlechten Behandlungsergebnis einher.

Die Untersuchung der *theoriegeleiteten Verlaufgruppen* erfolgte mit dem gleichen Prädiktorensatz und erbrachte hinsichtlich der Patienten- und Prozessmerkmale ähnliche Ergebnisse wie die explorativen Verlaufgruppen. Der zusätzliche Vorhersagewert der theoriegeleiteten Verlaufgruppen fällt dabei jedoch gering aus und erklärt keinen bedeutenden Varianzanteil im Gesamtmodell. Bei differenzierter Betrachtung zeigt sich für die Non-Response_{TG}-Gruppe ein Vorhersagewert des primären Outcomes: Der Verlauf einer Non-Response_{TG} geht mit einer erhöhten katamnestischen Symptombelastung, also einem schlechteren Outcome, einher. Die theoriegeleiteten Verlaufgruppen unterscheiden sich unter alleiniger Kontrolle der Ausgangsbelastung zwar bedeutend im primären Outcome, in den paarweisen Vergleichen zeigt sich jedoch, dass der Effekt durch den einzigen signifikanten Unterschied zwischen Early-Response_{TG}- und Non-Response_{TG}-Gruppe bedingt wird.

Der Vorhersagewert der Verlaufgruppen für das sekundäre Outcome, die katamnestisch erhobene *gesundheitsbezogene Lebensqualität*, fällt insgesamt deutlich schwächer aus. Hinsichtlich der *explorativen Verlaufgruppen* zeigen sich auch bei statistischer Kontrolle der Ausgangsbelastung bedeutende Unterschiede zwischen den Gruppen. Dabei weist der paarweise Vergleich der Gruppen Linear Response_{EG} und Early Response_{EG} ebenso wie der Vergleich der beiden Gruppen Delayed Response_{EG} und Non-Response_{EG} keine Unterschiede auf, die beiden Gruppenpaare unterscheiden sich hingegen deutlich voneinander. Die Gruppen Linear Response_{EG} und Early Response_{EG} weisen eine deutlich höhere katamnestische Lebensqualität auf als die Gruppen Delayed Response_{EG} und Non-Response_{EG}. Von den zur Vorhersage des sekundären Outcomes einbezogenen Patientenmerkmalen haben ein hoher Schulabschluss, eine bestehende Erwerbstätigkeit, eine positive Bewertung der Güte der therapeutischen Beziehung und ein hoher Ausgangswert der gesundheitsbezogenen Lebensqualität einen positiven prädiktiven Wert für die katamnestische Lebensqualität. Erhöhter Leidensdruck zu Behandlungsbeginn geht mit einer geringeren katamnestischen Lebensqualität einher. Durch den Einschluss der Prozessmerkmale Klinik, Behandlungsdauer sowie Dosis und Art psychotherapeutischer Maßnahmen erfolgt keine verbesserte Vorhersage des sekundären Outcomes. Die zusätzliche Varianzaufklärung durch die explorativen Verlaufgruppen fällt gering aber statistisch signifikant aus, was vor allem auf den Vorhersagewert der Non-Response_{EG}-Gruppen für eine geringere katamnestische Lebensqualität zurückzuführen ist.

Die *theoriegeleiteten Verlaufgruppen* unterscheiden sich unter Kontrolle des Ausgangswertes im sekundären Outcome *gesundheitsbezogene Lebensqualität* weder signifikant noch in der Höhe einer kleinen Effektstärke voneinander. Zur Untersuchung

möglicher prädiktiver Aussagen durch die theoriegeleiteten Verlaufgruppen wurde das sekundäre Outcome ebenfalls mit dem gleichen Set an Patienten- und Prozessmerkmalen wie zur Untersuchung der explorativen Verlaufgruppen analysiert. Dabei erwiesen sich die gleichen Patientenmerkmale (hoher Schulabschluss, Erwerbstätigkeit, positive Güte der therapeutischen Beziehung, hohe initiale Lebensqualität und geringer Leidensdruck) in weitgehend gleicher Ausprägung als bedeutsame Prädiktoren der katamnestischen Lebensqualität. Auch in diesem Modell erlangten die Prozessmerkmale Art, Dauer und Dosis der Behandlung keinen bedeutsamen Vorhersagewert. Für die fokussierten theoriegeleiteten Verlaufgruppen zeigte sich die Gruppenzugehörigkeit zur Late Response_{TG} tendenziell als Prädiktor für eine erhöhte katamnestische Lebensqualität. Im Gesamtmodell tragen die theoriegeleiteten Verlaufgruppen allerdings nur zu 1% an erklärter Varianz im Outcome bei, was sich zudem inferenzstatistisch nicht absichern lässt.

Für die *Hypothesen 4 bis 7*, die das primäre Outcome betreffen, können *Hypothese 4* und *Hypothese 5*, die für die explorativen Verlaufgruppen bedeutende Gruppenunterschiede sowie einen bedeutsamen Zuwachs an Varianzaufklärung im Outcome bei Berücksichtigung von Patienten- und Prozessmerkmalen postulieren, angenommen werden. Die *Hypothese 6*, in der ein bedeutender Gruppenunterschied zwischen den theoriegeleiteten Verlaufgruppen im primären Outcome angenommen wurde, wird beibehalten. Unter Berücksichtigung weiterer Patienten- und Prozessmerkmale tragen die theoriegeleiteten Verlaufgruppen jedoch keinen relevanten Zuwachs in der Varianzaufklärung des primären Outcomes bei. *Hypothese 7* konnte nicht angenommen werden.

Bei den *Hypothesen 8 bis 11*, die das sekundäre Outcome betreffen, zeigte sich in den explorativen Verlaufgruppen ein signifikanter Gruppenunterschied. Unter Berücksichtigung weiterer Patienten- und Prozessmerkmale tragen die explorativen Verlaufgruppen noch bedeutend zur Vorhersage des sekundären Outcomes bei. *Hypothese 8* und *Hypothese 9* werden angenommen. Die theoriegeleiteten Verlaufgruppen hingegen unterscheiden sich nicht bedeutend im sekundären Outcome. Bei Berücksichtigung weiterer konfundierender Merkmale (Patienten- und Prozessmerkmale) kann die Varianz im sekundären Outcome durch die theoriegeleiteten Verlaufgruppen nicht bedeutend erklärt werden. *Hypothesen 10* und *11* werden verworfen.

Zusammenfassend können aus den Ergebnissen der Hauptfragestellung einige grundlegende Schlussfolgerungen gezogen werden, die im Anschluss detailliert diskutiert werden:

- a) Die identifizierten Symptomverlaufgruppen unterscheiden sich bedeutend im primären Outcome, der katamnestischen Symptombelastung. Auch unter Berücksichtigung empirisch fundierter Merkmale ist das primäre Outcome durch die explorativen Symptomverlaufgruppen vorhersagbar. Patienten mit einer Early Response_{EG} profitieren zur Katamnese stärker, wohingegen die Zugehörigkeit zu den Gruppen Delayed_{EG} und Non-Response_{EG} ein schlechteres katamnestisches Outcome vorhersagt. Hinsichtlich des sekundären Outcomes, der katamnestischen gesundheitsbezogenen Lebensqualität, lassen sich nur für die explorativen Verlaufgruppen, jedoch nicht für die theoriegeleiteten Verlaufgruppen, bedeutsame Unterschiede finden. Werden weitere Prädiktoren einbezogen, ist das sekundäre Outcome ebenfalls nur über die explorativen Verlaufgruppen vorhersagbar. Der Anteil der erklärten Varianz ist jedoch geringer als im primären Outcome. Der Verlauf einer Non-Response_{EG} geht mit einer geringeren katamnestischen Lebensqualität einher. Generell erfolgt die Vorhersage des längerfristigen Behandlungsergebnisses besser über die explorativen als über die theoriegeleiteten Verlaufgruppen.
- b) Von den untersuchten Patientenmerkmalen stehen vor allem das Bildungsniveau, die Erwerbstätigkeit, die initiale Belastung und Merkmale, die mit chronifizierten psychischen Erkrankungen einhergehen in einem bedeutenden Zusammenhang mit dem längerfristigen Behandlungsergebnis. Ferner ist die Güte der therapeutischen Beziehung ein relevanter Prädiktor sowohl des primären als auch des sekundären Outcomes.
- c) Die Behandlungsdauer ist ein relevanter Prädiktor des primären Outcomes, der katamnestischen Symptombelastung. Die Dosis und die Art der psychotherapienahen Maßnahmen liefern unter Berücksichtigung aller anderen Merkmale keinen prädiktiven Vorhersagewert für die Varianz im primären und sekundären Outcome.

6.2.3.1 *Symptomverlauf als Prädiktor des Behandlungsergebnisses*

Die identifizierten Symptomverlaufgruppen erlauben eine Vorhersage der katamnestischen Symptombelastung. Patienten, die früh im Verlauf auf die Behandlung ansprechen und die Symptomreduktion im weiteren Verlauf weitgehend stabil erhalten, weisen ein besseres katamnestisches Behandlungsergebnis auf als alle anderen Verlaufgruppen. Die Veränderung der Symptombelastung von der Aufnahme bis zur Halbjahreskatamnese erreicht dabei in dieser Gruppe die Effektstärke von $d = 1.56$ für die explorative Early-

Response_{EG}-Gruppe und $d = 0.89$ für die theoriegeleitete Early-Response_{TG}-Gruppe. Effektstärken in dieser Größenordnung wurden in der stationären Behandlung psychisch erkrankter Patienten bislang nur im Prä-Post-Vergleich störungsspezifischer Outcome-Maße bei depressiven Patienten berichtet (vgl. Härter et al., 2004; Steffanowski et al., 2007). Die erreichte Prä-Katamnese-Effektstärke der Gesamtstichprobe liegt mit $d = 0.54$ im erwarteten Bereich. Steffanowski et al. (2007) berichten für den Prä-Katamnese-Vergleich über 46 Studien eine Gesamteffektstärke von $d = 0.41$ bei einem durchschnittlichen Katamnese-Zeitraum von 12 Monaten. Für die explorativen Verlaufsgruppen erreichen die Patienten der Early-Response_{EG}-Gruppe mit Abstand die höchste Effektstärke. Interessant stellen sich auch die Ergebnisse der explorativen Non-Response_{EG} und Linear Response_{EG} dar: Während die Linear-Response_{EG}-Gruppe zum Ende der Behandlung erstmal einen Abfall und im späteren Verlauf bis zur Katamnese wieder einen leichten Anstieg der Symptombelastung aufweist, zeigen die Non-Response_{EG}-Patienten ein umgekehrtes Muster. Bis zum Ende der Behandlung lässt sich keine bedeutende Verbesserung nachweisen, zur Katamnese ist jedoch ein Rückgang der Symptombelastung zu verzeichnen. Hiermit lassen sich auch die relativ ähnlichen Prä-Katamnese-Effektstärken der beiden Gruppen mit $d = 0.33$ und $d = 0.43$ erklären. Der Verlauf der Linear-Response_{EG}-Gruppe spiegelt den typischen Prä-Post-Katamnese-Verlauf wider, der häufig in der stationären psychotherapeutischen Behandlung gefunden wird. Da die Gruppe die größte Verlaufsgruppe bildet, kann dieses Ergebnis als Bestätigung bisheriger Befunde gesehen werden. Der Verlauf der Non-Response_{EG}-Gruppe ist dabei eher ungewöhnlich. Eine denkbare Erklärung wäre, dass bei diesen Patienten andere therapeutische Mechanismen greifen, die erst verzögert zu einer Symptomentlastung führen. Dies könnte im Sinne psychodynamischer Therapiekonzepte verstanden werden, in denen es nicht vorrangig um die Behandlung der Symptomatik geht, sondern vor allem um die Verarbeitung unbewusster seelischer Konflikte³⁷ (Streeck, 2005). Zur Festigung dieser Annahme müssten allerdings die angewandten therapeutischen Verfahren differenzierter untersucht werden. Die anteilige Dosis psychodynamischer Verfahren³⁸ liegt zwar in der Non-Response_{EG}- ebenso wie in der Delayed-Response_{EG}-Gruppe höher als in der Early-Response_{EG}- und Linear-Response_{EG}-Gruppe, die Unterschiede zwischen den Gruppen erweisen sich jedoch als nicht signifikant (s. Anhang: Tabelle C.3). In den Vorhersagemodellen des Outcomes konnte die therapieschulenspezifische Behandlungsdosis³⁹ ebenfalls keinen bedeutenden Beitrag zur Erklärung der Varianz leisten.

³⁷ Die Angemessenheit der eingesetzten Ergebniskriterien wird gesondert in Abschnitt 6.3.6 diskutiert.

³⁸ Der Anteil der Dosis psychodynamischer Therapiemaßnahmen wurde an der Gesamtdosis der psychodynamischen und verhaltenstherapeutischen Behandlungsmaßnahmen relativiert.

³⁹ In der vorliegenden Arbeit als *Art der psychotherapienahen Behandlung* beschrieben.

Ein anderer Ansatz zur Erklärung des Verlaufs der Non-Response_{EG}-Gruppe liegt in der Vermutung, dass diese Patienten eine längere Zeit brauchen, um eine therapeutische Arbeitsbeziehung aufzubauen. Folglich könnte eine positive und effektive therapeutische Beziehung erst gegen Ende der Behandlung entstanden sein, was dann zu einem verzögerten Auftreten der Symptomreduktion führen kann. Diese Annahme wird durch die Ergebnisse der 2. Fragestellung gestützt, in der die Güte der therapeutischen Beziehung in der Non-Response_{EG}-Gruppe am schlechtesten eingeschätzt wurde. Eine weitere Erklärung könnte jedoch auch beinhalten, dass die Patienten dieser Gruppe tatsächlich wenig bis gar nicht von der stationären Behandlung profitieren und die zur Katamnese auftretenden Effektstärken auf methodische Artefakte der Spontanremission und der Regression zurückzuführen sind. Auf beide Konzepte wird im Abschnitt 6.3.2 der methodischen Diskussion vertiefend eingegangen.

Die Delayed-Response_{EG}-Gruppe weist einen auffälligen Verlauf auf, bei dem ein starker Rückgang der Symptombelastung erst gegen Ende der Behandlung auftritt, dieser jedoch zur Katamnese wieder zurückfällt. Zwar erreichen die Patienten dieser Gruppe im Mittel Veränderungen einer mittleren Effektstärke ($d = 0.54$), betrachtet man die Veränderung von der Aufnahme bis zur Entlassung (vgl. Abbildung 26, S. 157), fällt auf, dass hier ein deutlich größerer Fortschritt erzielt wurde. Für diese Patienten ist die Ableitung von klinischen Implikationen besonders bedeutend. Anscheinend werden erst in der letzten Phase der Behandlung Verbesserungen erzielt, die jedoch nicht lange aufrechterhalten werden können. Hier stellt sich die Frage, ob diese Patienten ausreichend lange behandelt werden. Möglicherweise brauchen diese Patienten eine längere Behandlung, um die erzielten Behandlungseffekte auch längerfristig stabilisieren zu können. Andererseits könnte die Dauer der *stationären* Behandlung zwar ausreichend, eine gute Integration der Veränderungsmechanismen in den Alltag jedoch erschwert sein. Entsprechend wäre für diese Patientengruppe, die zu Behandlungsbeginn am höchsten belastet ist, eine verbesserte Nachsorgebehandlung indiziert. Diese könnte über eine verstärkte Fokussierung der Nachsorgeplanung schon während des Klinikaufenthaltes durch den Therapeuten vorbereitet werden.

Zusammenfassend zeigen alle der explorativen Verlaufgruppen eine Verbesserung der Symptombelastung mit mindestens kleiner Effektstärke. Dabei scheint die Einteilung in die Verlaufgruppen vor allem eine Differenzierung der Behandlungsergebnisse zwischen den Patientengruppen mit hoher initialer Symptombelastung zu ermöglichen.

Bei der theoriegeleiteten Gruppeneinteilung findet sich ein anderes Bild: Insgesamt variieren die Effektstärken weniger stark zwischen den Gruppen als bei den explorativen Verlaufgruppen. Im Hinblick auf die Symptombelastung erreicht die Early-Response_{TG}-Gruppe eine mittlere Prä-Katamnese-Effektstärke von $d = 0.89$. Die Late-Response_{TG}-

Gruppe erreicht eine mittlere Effektstärke von $d = 0.60$. Lediglich die Non-Response_{TG}-Gruppe bleibt bis zur Katamnese auf dem gleichen Belastungsniveau wie zu Behandlungsbeginn und zur Entlassung ($d = 0.06$). Allerdings ist diese Gruppe auch diejenige, die zu Behandlungsbeginn die geringste Symptombelastung aufweist. Betrachtet man die Werte der Symptombelastung zum Katamnesezeitpunkt, wird deutlich, dass die Gruppenmittelwerte alle sehr eng um den Gesamtmittelwert streuen: Wie auch die Ergebnisse der linearen Regression zeigen, tragen die theoriegeleiteten Verlaufgruppen keinen bedeutenden Beitrag zur Varianzaufklärung im längerfristigen Behandlungsergebnis bei.

Für das sekundäre Outcome stellen sich die Ergebnisse ähnlich dar. Die explorativen Verlaufgruppen ermöglichen eine Vorhersage der Varianz in der gesundheitsbezogenen Lebensqualität, selbst wenn fundierte Patienten- und Prozessmerkmale berücksichtigt werden. Allerdings ist der inkrementelle Beitrag der erklärten Varianz nur noch von mittlerer Effektstärke. Die theoriegeleiteten Verlaufgruppen unterscheiden sich dahingegen nicht bedeutend in der katamnestischen Lebensqualität. Auch unter Einschluss weiterer Merkmale tragen die theoriegeleiteten Verlaufgruppen nicht bedeutend zur erklärten Varianz im sekundären Outcome bei. Demzufolge kann auch kein mediiender Effekt über die hier untersuchten Variablen angenommen werden. Für beide Ergebniskriterien lassen sich ähnliche Patientenmerkmale finden, die im Folgenden detailliert diskutiert werden. Anders als bei der Vorhersage des primären Outcomes wurde beim sekundären Outcome auch der Ausgangswert des Outcome-Kriteriums ebenso wie die FPTM-Skala *Leidensdruck* einbezogen. Es zeigt sich, dass unter anderem eine hohe initiale Lebensqualität und ein geringer Leidensdruck die katamnestische Lebensqualität bedeutend vorhersagen. Dennoch tragen die explorativen Verlaufgruppen Delayed Response_{EG} und Non-Response_{EG} (im Vergleich zur Referenzgruppe Linear Response_{EG}) bedeutend negativ zur Vorhersage der katamnestischen Lebensqualität bei. Da die Early-Response_{EG}-Gruppe hingegen keinen bedeutenden Vorhersagewert leistet, kann dies nicht über die vergleichsweise hohe Ausgangsbelastung der drei Gruppen erklärt werden. Patienten die spät im Verlauf ansprechen, ebenso wie Patienten, deren Symptombelastung sich während der Behandlung nicht wesentlich ändert, weisen katamnestisch eher eine niedrige Lebensqualität auf, als die initial auch hoch belasteten Patienten, die sich aber früh im Behandlungsverlauf verbessern. Die Einteilung in Verlaufgruppen über theoriegeleitete Kriterien erlaubt keine Vorhersage der katamnestischen Lebensqualität.

Erklärungsansätze für die unterschiedlichen Ergebnisse, die aus den Vorhersagen über die explorativen und die theoriegeleiteten Verlaufgruppen hervorgehen, beziehen sich hauptsächlich auf methodische Aspekte der Gruppenbildung und werden im Abschnitt 6.2.3.4 gesondert diskutiert.

6.2.3.2 Patientenmerkmale als Prädiktoren des Behandlungsergebnisses

Die *Schulbildung* ebenso wie der *Erwerbsstatus* stellen bedeutende Prädiktoren des katamnestic Outcomes dar. Dieser Zusammenhang lässt sich konsistent über die beiden untersuchten Outcome-Maße finden und geht einher mit den schon in den Prädiktoranalysen gefundenen Zusammenhängen: Ein niedriges Bildungsniveau und eine bestehende Erwerbslosigkeit weisen eine erhöhte Wahrscheinlichkeit für einen ungünstigen Verlauf sowohl während der Behandlung (vgl. Abschnitt 6.2.2.1) als auch hinsichtlich längerfristiger Behandlungsergebnisse auf. Ein niedriger sozioökonomischer Status scheint ein relevanter Prädiktor für ungünstige Prozesse und die längerfristige Wirksamkeit von Psychotherapie sein. Dabei wird der Zusammenhang zwischen sozioökonomischem Status und längerfristigem Outcome nicht alleine durch den ungünstigen Verlauf bedingt, der prädiktive Wert bleibt auch unter Berücksichtigung des Verlaufs erhalten. Ebenso kann aus den vorliegenden Ergebnissen die Annahme eines mediierenden Effektes der Dosis von psychotherapienahen Maßnahmen nicht bestätigt werden. Auch unter Einbezug der Dosis blieb das Bildungsniveau ein stabiler Prädiktor für ein schlechteres Outcome. Zumindest in dieser Stichprobe scheint das schlechtere Behandlungsergebnis nicht durch den erschwerten Zugang zur Psychotherapie innerhalb der stationären Behandlung bedingt zu sein. Vielmehr zeigte sich, dass die Dosis psychotherapienaher Maßnahmen, unter Berücksichtigung weiterer Prozess- und Patientenmerkmale, keinen bedeutenden Zusammenhang mit dem Outcome aufwies⁴⁰. Insgesamt verdeutlichen die Ergebnisse den hohen Stellenwert der Merkmale Bildungsniveau und Erwerbsstatus für die Wirksamkeit stationärer psychotherapeutischer Behandlungen. Einschränkend muss allerdings die Operationalisierung der Merkmale Schulbildung und Erwerbsstatus erwähnt werden. Beide Merkmale gingen in den Analysen als dichotome Variablen ein, sodass es sich um eine sehr grobe Einteilung handelt: Die Kategorie hohes Bildungsniveau zeichnet sich durch den mindestens erreichten Realschulabschluss aus. Die Erwerbslosigkeit umfasste sowohl nicht erwerbstätige als auch erwerbslose Patienten. Beide Einteilungen können aufgrund der reduzierten Information kritisch betrachtet werden. Eine weiter ausdifferenzierte Einteilung hätte hier allerdings zu zum Teil sehr kleinen Untergruppen innerhalb der Verlaufsgruppen geführt (s. Anhang: Tabelle C.1 und Tabelle D.1), die nicht mit ausreichender Teststärke hätten analysiert werden könnten. Die vorgenommene Einteilung stellt demnach eine gute Annäherung dar, die auch in anderen Studien in ähnlicher Weise verwendet wird (Steffanowski et al., 2007).

⁴⁰ Auf die Diskussion des Zusammenhangs von Dosis und Dauer mit dem Behandlungsergebnis wird in Abschnitt 6.2.3.3 gesondert eingegangen.

Ein unerwartetes Ergebnis stellt der *Vorhersagewert der Diagnose* für das Behandlungsergebnis dar. In den multivariaten Analysen zeigte sich, dass die Diagnose unter Berücksichtigung weiterer Patienten- und Prozessmerkmale fast keinen Vorhersagewert für das katamnestic Outcome hat. Lediglich die Diagnose einer Anpassungsstörung trägt bedeutend zur Varianzaufklärung im Outcome bei. In bisherigen Studien aus dem stationären Setting wurden indikationsspezifische Unterschiede in den erreichten Effektstärken gefunden. So weisen Patienten mit somatoformen Störungen die geringsten Effektstärken und Patienten mit depressiven Störungen die höchsten Effektstärken auf (Barghaan et al., 2005; Steffanowski et al., 2007). Die hier erzielten Ergebnisse liefern Hinweise darauf, dass die Diagnose möglicherweise gar keine so große Rolle für die Wirksamkeit der Behandlung spielt, wenn weitere Prozessmerkmale wie der Verlauf der Symptombelastung, die Behandlungsdauer und die Güte der therapeutischen Beziehung berücksichtigt werden. Es scheint also nicht vorrangig von der zugrundeliegenden psychischen Störung abzuhängen, ob die Patienten längerfristig von der Behandlung profitieren, sondern vielmehr davon, wie sich die Prozesse während der stationären Behandlung gestalten. Kommt es im frühen Verlauf der Behandlung zu einer guten Arbeitsbeziehung zwischen Patient und Therapeut, und wird diese Arbeitsbeziehung durch eine Verbesserung der Symptomatik begleitet⁴¹, so können die Behandlungseffekte auch poststationär über einen längeren Zeitraum erhalten werden. Die Symptomverlaufsgruppen zeichnen sich zwar durch eine erhöhte Wahrscheinlichkeit für bestimmte Diagnosegruppen aus (vgl. Abschnitt 6.2.2.1), hinsichtlich des längerfristigen Behandlungsergebnisses tragen die Diagnosen jedoch fast gar nicht mehr zur Varianzaufklärung bei. Der Zusammenhang zwischen dem Verlauf der Symptombelastung und dem Behandlungsergebnis scheint den Zusammenhang von Diagnose und Outcome fast vollständig zu erklären. Ausschließlich die Diagnose einer Anpassungsstörung trägt noch einen bedeutenden Anteil zur Varianz im Behandlungsergebnis bei, wenn weitere Patienten- und Prozessmerkmale berücksichtigt werden. Eine Anpassungsstörung steht im Zusammenhang mit einer geringeren Symptombelastung zur Katamnese. Die Anpassungsstörung ist eine nicht-chronifizierte Erkrankung und zeichnet sich unter anderem durch eine geringer ausgeprägte Symptomschwere als die unipolare Depression aus (Kriterien nach ICD-10; Dilling et al., 2000). Entsprechend lassen sich die Ergebnisse dahingehend interpretieren, dass möglicherweise die Erkrankungsdauer und die Chronifizierung der Erkrankung im Zusammenhang zum längerfristigen Behandlungsergebnis stehen. Dies würde auch den meta-analytischen Befunden entsprechen, in denen die Erkrankungsdauer den stärksten Prädiktor für das Behandlungsergebnis darstellt (Steffanowski et al., 2007). Die

⁴¹ Auf das Problem der Kausalität der beiden Prozessmerkmale Güte der therapeutischen Beziehung und Symptombelastung wurde in Abschnitt 6.2.2.1 eingegangen.

Erkrankungsdauer als Maß der Chronifizierung wurde in dieser Studie nicht erhoben, sodass diese Annahme in den vorliegenden Daten nicht überprüft werden konnte. In Folgestudien sollte die Untersuchung der Chronifizierung stärker fokussiert werden. Methodische Einschränkungen hinsichtlich der Ergebnisse zu den Diagnosegruppen werden in Abschnitt 6.3.4.1 diskutiert.

Entgegen der Erwartung durch bisherige Befunde trägt die *Psychotherapiemotivation* weder zur Vorhersage des primären noch des sekundären Outcomes bei. In anderen Untersuchungen, in denen das gleiche Instrument zur Erfassung der Psychotherapiemotivation im stationären Setting eingesetzt wurde, finden sich signifikante Zusammenhänge mit dem Behandlungsergebnis (Barghaan et al., 2005; Nübling, Schmidt & Schulz, 2002; Nübling et al., 2006). Die Ausprägung des Zusammenhangs betrug jedoch zur 1-Jahres-Katamnese nur noch für die Skala *Hoffnung* eine mittlere Effektstärke von $r = .33$, das multiple R über alle sechs Skalen lag bei $.41$ (Nübling et al., 2006). Anders als in bisherigen Studien wurde in der vorliegenden Untersuchung eine Vielzahl weiterer Patientenmerkmale zur Prädiktion des Behandlungsergebnisses einbezogen. Dementsprechend könnte beispielsweise die Güte der therapeutischen Beziehung, die hier einen bedeutenden Vorhersagewert für das Outcome aufweist, als Mediator zwischen der Psychotherapiemotivation und dem Behandlungsergebnis fungieren: Vorstellbar wäre, dass eine hohe initiale Psychotherapiemotivation eine gute Beziehungsgestaltung zwischen Patient und Therapeut ermöglicht, die längerfristig ein positives Behandlungsergebnis begünstigt. Da in bisherigen Studien nicht beide Konstrukte gemessen wurden, und der empirisch belegte Zusammenhang zwischen Psychotherapiemotivation und Behandlungsergebnis hier unter Einbezug der therapeutischen Beziehung nicht aufgetreten ist, könnte eine vollständige Mediation durch die therapeutische Beziehung gegeben sein.

Die *Güte der therapeutischen Beziehung* zeigt sich in den Outcome-Analysen durchgängig als relevanter Prädiktor des längerfristigen Behandlungsergebnisses. Sowohl im primären als auch im sekundären Outcome geht eine positive Einschätzung der therapeutischen Beziehung mit einem besseren katamnestischen Behandlungsergebnis einher. Diese Ergebnisse sind konsistent mit den positiven empirischen Befunden zur Vorhersage der Wirksamkeit von ambulanter Psychotherapie (Beutler et al., 2004; Martin et al., 2000; Orlinsky et al., 2004). Für die stationäre Behandlung liegen weniger Studien vor, die ebenfalls auf einen positiven Zusammenhang der therapeutischen Beziehung mit dem Behandlungsergebnis hindeuten (Bassler et al., 1995b; Blais, Jacobo & Smith, 2010; Sharf et al., 2010). Die Ergebnisse dieser Arbeit zeigen eindrücklich, dass die Güte der therapeutischen Beziehung auch bei gleichzeitiger Untersuchung weiterer relevanter

Prädiktoren mit einer geringeren katamnestischen Symptombelastung und erhöhten Lebensqualität (psychische Aspekte der gesundheitsbezogenen Lebensqualität) zusammenhängt.

Um so erstaunlicher ist das bereits in Abschnitt 6.2.2.1 diskutierte Ergebnis des geringen Zusammenhangs zwischen der Güte der therapeutischen Beziehung und dem Symptomverlauf während der Behandlung: Lediglich die Zugehörigkeit zu der theoriegeleiteten Non-Response_{TG}-Gruppe kann erwartungskonform durch eine negativere Einschätzung der therapeutischen Beziehung vorhergesagt werden. Zur Vorhersage der Symptomverlaufgruppen wurden zwar mehrere Prädiktoren untersucht, sodass mediierende oder Interaktionseffekte mit anderen Merkmalen den Zusammenhang verdecken könnten, allerdings wurden in die Analysen zur Vorhersage des Behandlungsergebnisses die gleichen Prädiktormerkmale berücksichtigt. Eine andere Erklärung könnte die Nachhaltigkeit des Zusammenhangs betreffen: Eine gute therapeutische Beziehung zu Beginn der Behandlung hat möglicherweise weniger Bedeutung für den zeitnahen Verlauf der Symptombelastung als für die Nachhaltigkeit der erreichten Behandlungsergebnisse. Die Schwierigkeit der Kausalbeziehung von therapeutischer Beziehung und Symptomverlauf wurde bereits in Abschnitt 6.2.2.1 diskutiert. Für den Zusammenhang zum längerfristigen Outcome scheint die Richtung des Zusammenhangs eindeutiger auszufallen, wobei in Eingruppen-Prä-Post-Designs generell keine Kausalbeziehungen abgeleitet werden können, da der Einfluss (nicht untersuchter) Confounder nicht ausgeschlossen werden kann (vgl. Abschnitt 2.2).

Insgesamt können die Ergebnisse zum Zusammenhang der Güte der therapeutischen Beziehung mit dem längerfristigen Behandlungsergebnis richtungsweisend für weitere Untersuchungen zu den Wirkfaktoren stationärer Psychotherapie genutzt werden. Für die klinische Praxis kann die Beforschung der therapeutischen Beziehung wichtige Erkenntnisse zur Verbesserung und Verstetigung des Behandlungserfolgs liefern.

6.2.3.3 Dauer und Dosis als Prädiktoren des Behandlungsergebnisses

Die Frage nach dem Zusammenhang von Dauer und Wirksamkeit der stationären Behandlung ist vor dem Hintergrund der Diskussion um die Kosten-Nutzen-Relation und die gerechte und adäquate Verteilung von Ressourcen vor allem für die Versorgungspraxis von hoher Relevanz (Lutz et al., 2003; Nosper, 2008; Spießl & Klein, 2008). Die bisherigen Forschungsergebnisse weisen jedoch ein uneinheitliches Bild auf (Hölzel et al., 2010; Kösters et al., 2006; Nosper, 1999a; Schmitz-Buhl et al., 1999; Steffanowski et al., 2007; Zielke, 2006). Erklärungen hierfür wie die Heterogenität der untersuchten Patientenstichproben und der Versorgungssysteme, die eingeschränkte Varianz der Behandlungsdauer aber auch die generelle Schwierigkeit, diese Frage in Beobachtungsstudien zu untersuchen, wurden bereits in Abschnitt 2.5 erläutert.

In den Ergebnissen der vorliegenden Arbeit stellt die Behandlungsdauer unter Einbezug weiterer Patienten- und Prozessmerkmale einen relevanten Prädiktor des primären Outcomes, d.h. der katamnestischen Symptombelastung, dar. Für das sekundäre Outcome, die katamnestische gesundheitsbezogene Lebensqualität, lässt sich dieser Zusammenhang nicht nachweisen. Die Dosis und die Art der psychotherapienahen Maßnahmen liefern unter Berücksichtigung aller anderen Merkmale keinen prädiktiven Vorhersagewert für die Varianz im primären und sekundären Outcome. Betrachtet man die bivariaten Zusammenhänge zwischen Dauer bzw. Dosis und primärem und sekundärem Outcome, fällt auf, dass die Behandlungsdauer in keinem bedeutenden Zusammenhang mit der katamnestischen Symptombelastung steht ($r = .052$; s. Anhang: Tabelle A.4). Wird die Dauer der Behandlung aber unter Berücksichtigung weiterer Patienten- und Prozessmerkmale, einschließlich der explorativen Verlaufsgruppen, untersucht, erlangt diese einen bedeutenden prädiktiven Wert für das katamnestische primäre Outcome: Längere Behandlungen gehen mit einer geringeren Symptombelastung 6 Monate nach Beendigung der stationären Behandlung einher. Dies legt die Annahme nahe, dass der Zusammenhang zwischen Behandlungsdauer und -outcome über eine oder mehrere Drittvariablen moderiert wird. Neben der Behandlungsdauer trugen ein hoher Schulabschluss, die Erwerbstätigkeit, die Diagnose einer Anpassungsstörung, eine positive Beurteilung der Güte der therapeutischen Beziehung und der Verlauf einer Early Response bedeutend zur Vorhersage der Varianz im primären Outcome bei. Entsprechend könnte eine längere Behandlung *nur dann* einen positiven Einfluss auf das längerfristige Behandlungsergebnis haben, wenn eine gute therapeutische Beziehung besteht und/oder schon früh während der Behandlung eine Besserung der Symptombelastung auftritt. Diese Annahme würde für die klinische Praxis implizieren, dass Patienten mit dem Verlauf einer Early Response *gerade nicht* früher entlassen werden sollten, da vor allem bei diesen Patienten ein positiver Effekt einer längeren Behandlung zu erwarten wäre. Denkbar wären aber auch andere Moderatoreffekte wie eine höhere Schulbildung, die einen positiven Zusammenhang zwischen der Dauer und dem Outcome (mit-)bedingen könnte. Dies könnte im Zusammenhang mit den Ergebnissen von Hölzel et al. (2010) wie folgt interpretiert werden: Patienten mit höherer Schulbildung weisen nicht *per se* ein besseres Behandlungsergebnis auf, sondern erhalten eher psychotherapeutische Behandlungen bzw. werden eher länger behandelt, was zu einem besseren Outcome führt. Die Dosis der psychotherapeutischen Behandlungsmaßnahmen trug in den vorliegenden Ergebnissen jedoch nicht bedeutend zur Aufklärung der Varianz im Outcome bei, sodass für die vorliegenden Daten eher die Dosis aller Behandlungen (äquivalent zur Behandlungsdauer) von Relevanz zu sein scheint. Allerdings ist das von Hölzel et al. (2010) untersuchte psychiatrische Setting, in dem die psychotherapeutische Behandlung nicht den gleichen hohen Stellenwert einnimmt wie in dem hier untersuchten Setting, und der Zeitpunkt

der Ergebnismessung (bei Entlassung) nur eingeschränkt vergleichbar mit der vorliegenden Untersuchung.

Insgesamt können über potentiell vermittelnde Effekte nur Vermutungen angestellt werden, da keine Interaktionseffekte zwischen den einzelnen Prädiktoren berechnet wurden. Für Folgeuntersuchungen wäre eine verstärkte Untersuchung der Interaktion zwischen Prozess- und Patientenmerkmalen empfehlenswert.

Die Varianz im sekundären Outcome konnte nicht über die Behandlungsdauer vorhergesagt werden. Hier stellen die Merkmale hoher Schulabschluss, Erwerbstätigkeit, eine positive therapeutische Beziehung, geringer initialer Leidensdruck und eine hohe initiale Lebensqualität relevante Prädiktoren für eine hohe katamnestische Lebensqualität dar. Der Verlauf mit einer Delayed Response_{EG} bzw. Late Response_{TG} oder einer Non-Response_{EG} nach der explorativen Gruppeneinteilung geht hingegen eher mit einer geringeren katamnestischen Lebensqualität einher. Hier wäre ebenso eine Untersuchung potentieller Interaktionen zwischen der therapeutischen Beziehung, dem Symptomverlauf und der Dauer der Behandlung denkbar. Generell muss die eingeschränkte Varianz der Behandlungsdauer berücksichtigt werden. Die Hälfte der Patienten der Verlaufsstichprobe wurde 6 Wochen behandelt, über 70% wiesen eine Behandlungsdauer zwischen 5 und 7 Wochen auf (s. Anhang: Tabelle C.3 und Tabelle D.3). Dies ist zum einen auf den hier vorgenommenen Ausschluss von Behandlungen unter 3 und über 9 Wochen zurückzuführen. Zum Anderen spiegelt dieses Spektrum an Behandlungszeiten die Versorgungsrealität wider (Deutsche Rentenversicherung Bund, 2010; Verband Deutscher Rentenversicherungsträger, 2004; vgl. auch Abschnitt 2.1.2), was die Untersuchung von Dauer-Wirkungs-Zusammenhängen erschwert (Kösters et al., 2006; Richter, 2001; Spießl & Klein, 2008).

Die Dosis und die Art der psychotherapienahen Maßnahmen tragen keinen wesentlichen Anteil an der Aufklärung der Varianz im primären und sekundären Outcome bei. Auch bivariat lassen sich keine Zusammenhänge zwischen der Art sowie Dosis und den Outcome-Maßen nachweisen. Dahingegen geht eine höhere Dosis psychotherapienaher Behandlungsmaßnahmen mit einer längeren Behandlung einher ($r = .38$; s. Anhang: Tabelle A.4), was für das angenommene proportionale Verhältnis von Dosis und Dauer der stationären Behandlung spricht (Paar & Grohmann, 2000; Steffanowski et al., 2007; vgl. auch Abschnitt 2.5.1). Dieses Ergebnis kann im Sinne des multimodalen Behandlungskonzepts erklärt werden: Für die Vorhersage der Wirksamkeit der stationären Behandlung scheint es schwierig zu sein, das konzeptuell so angelegte komplexe Behandlungsgeschehen in einzelne Behandlungseinheiten aufzugliedern. Die Ergebnisse legen nahe, stationäre Behandlungen als Gesamtheit zu untersuchen. Darüber hinaus ist die Validität der Dosisangaben über die KTL-Kodierung kritisch zu betrachten, die im Einzelnen in Abschnitt 6.3.4.2 diskutiert wird.

Zusammenfassend lässt sich festhalten, dass die Dauer der Behandlung einen relevanten positiven Prädiktor der katamnestischen Symptombelastung darstellt, wenn die Patientenmerkmale Bildung, Erwerbstätigkeit, die Diagnose einer Anpassungsstörung, die Güte der therapeutischen Beziehung und der Symptomverlauf während der Behandlung berücksichtigt werden. Die Dosis und die Art der psychotherapienahen Maßnahmen stehen in keinem bedeutenden Zusammenhang zum Outcome. Von besonderer Relevanz ist das Ergebnis, dass durch den Verlauf der Symptombelastung während der Behandlung, zu den genannten Patientenmerkmalen und der Behandlungsdauer, noch ein deutlicher Zuwachs an Varianzaufklärung im Outcome ermöglicht wird. Nicht nur die Ausprägung bestimmter relevanter Patientenmerkmale und die Dauer der Behandlung stehen im Zusammenhang mit dem längerfristigen Behandlungsergebnis, ein günstiger Behandlungsprozess im Sinne des Symptomverlaufs mit einer Early Response_{EG} trägt zusätzlich noch wesentlich zur Wirksamkeit der Behandlung bei, während der Verlauf einer Delayed Response_{EG} oder einer Non-Response_{EG} mit einem schlechteren Behandlungsergebnis einhergehen.

6.2.3.4 Vergleich der explorativen und theoriegeleiteten Verlaufgruppen zur Vorhersage des längerfristigen Behandlungsergebnisses

Die Analysen zur Vorhersage des katamnestischen Behandlungsergebnisses zeigten für das primäre Outcome eine deutliche Überlegenheit der explorativen Verlaufgruppen gegenüber den theoriegeleiteten Verlaufgruppen. Auch das sekundäre Outcome konnte durch die explorativen Verlaufgruppen noch signifikant vorhergesagt werden, während die theoriegeleiteten Verlaufgruppen keine Vorhersage ermöglichten.

Die Überlegenheit der explorativen Verlaufgruppen hinsichtlich der Vorhersage des längerfristigen Outcomes ist jedoch unter der Einschränkung zu interpretieren, dass durch diese Gruppeneinteilung nur über einen relativ kleinen Anteil der Patientenstichprobe, nämlich 29%, Aussagen gemacht werden können. Die Gruppe der Linear Response_{EG} umfasst 71% der untersuchten Stichprobe. Für diese Patienten kann nach der explorativen Vorgehensweise keine weitere Differenzierung im Behandlungsergebnis vorgenommen werden. Die Vorhersage des Behandlungsergebnisses über die theoriegeleiteten Verlaufgruppen ist in den vorliegenden Daten nur bedingt und in einem geringen Ausmaß möglich, was auf eine stärkere Homogenität hinsichtlich der initialen Symptombelastung zwischen den Gruppen zurückgeführt werden kann. Dies wird auch darin deutlich, dass der größte Anteil an erklärter Varianz in der katamnestischen Symptombelastung auf die initiale Symptombelastung zurückgeht. In die Bildung der theoriegeleiteten Verlaufgruppen ist die initiale Belastung nur indirekt eingegangen. Die Ergebnisse der explorativen Verlaufgruppen ebenso wie bisherige Befunde zur Vorhersage des psychotherapeutischen Behandlungserfolgs zeigen jedoch, welchen hohen Stellenwert die Ausprägung der initialen

Symptomschwere einnimmt (Beutler et al., 2006; Clarkin & Levy, 2004; Härter et al., 2004; Steffanowski et al., 2007; vgl. Abschnitt 2.4.1). Neben der initialen Symptombelastung ist aber auch der Verlauf über den gesamten Behandlungszeitraum von Relevanz. Während drei der explorativen Verlaufgruppen eine ähnlich hohe initiale Symptombelastung aufweisen, zeigen sich zwischen diesen Gruppen deutliche Unterschiede im weiteren Verlauf der Behandlung ebenso wie in den katamnestischen Ergebniskriterien. Die zeitlich punktuelle Betrachtung der Symptomverbesserung, wie sie der Bildung der theoriegeleiteten Gruppen zu Grunde liegt, scheint demnach kein ausreichendes Kriterium für die Bildung distinkter Subgruppen, die sich auch im Behandlungsergebnis unterscheiden, zu sein. Wie ist dieses Ergebnis im Kontext der bisherigen Befunde zum Phänomen der Early Response zu verstehen? Wie Lambert (2005) in seinem Übersichtsartikel zusammenfasst, zeigte sich in der Forschung der letzten drei Jahrzehnte eine deutliche Überlegenheit dieser Gruppe hinsichtlich des Behandlungsergebnisses. Alle berichteten Studien beziehen sich allerdings auf das ambulante Setting. Auch wurde schon auf die Uneinheitlichkeit des verwendeten zeitlichen und inhaltlichen Kriteriums einer Response hingewiesen. Zur Berücksichtigung anderer inhaltlicher Response-Kriterien, wurden die Verbesserungsdaten auch nach drei weiteren gängigen Definitionen in Fragestellung 1 D dargestellt. Es zeigte sich, dass die jeweilige wöchentliche Anzahl der verbesserten Patienten zwischen den Kriterien variiert: Das Kriterium einer Verbesserung mit mindestens mittlerer Effektstärke erreichten mehr Patienten zu einem frühen Zeitpunkt während der Behandlung (zum Beispiel nach der dritten Behandlungswoche 55%) als das Kriterium der 50%igen Reduktion der Symptombelastung (41%) oder der klinische signifikante Verbesserung (36%; vgl. Abschnitt 5.5). Die Ergebnisse verdeutlichen, wie stark der Anteil an Early Response-Patienten, und dementsprechend möglicherweise auch der Zusammenhang mit dem Behandlungsergebnis, von der Wahl des Verbesserungskriteriums abhängt. Ähnlich verhält es sich bei der Festlegung des zeitlichen Kriteriums: Aufgrund der geringen Anzahl von Studien zum Verlauf in der stationären Behandlung basiert der gewählte Zeitpunkt lediglich auf inhaltlichen Überlegungen, ein anderer Zeitpunkt oder ein zur Behandlungsdauer relativ festgelegter Zeitpunkt wäre inhaltlich auch ableitbar. In der hier gewählten Operationalisierung sollte dieser Unsicherheit Rechnung getragen werden, indem nicht nur ein festgelegter Zeitpunkt, sondern zwei Zeitpunkte in die Definition einer Response eingehen. Dies kann jedoch auch mit einer erhöhten Unschärfe und dem Risiko, auch nur kurzfristig auftretende Verbesserungen als Response zu definieren, einhergehen. Denkbar wäre eine Operationalisierung, bei der eine gewisse Kontinuität der Verbesserung über die ersten x Wochen berücksichtigt würde. Dies würde jedoch erneut die wöchentlichen Erhebungen, zumindest für einen Teil der Behandlungszeit, erfordern. Dieses Vorgehen wurde von Stulz et al. (2007) angewendet, indem die Autoren für die ersten 6 Wochen der Behandlung GMM-

Analysen durchgeführt haben. Hier konnten fünf distinkte Verlaufgruppen, die einen Zusammenhang zum Outcome aufwiesen, identifiziert werden. Die Frage, ab wann von einer „frühen“ Verbesserung gesprochen wird, wird auch sehr unterschiedlich behandelt. Aufgrund der stark variierenden Behandlungszeiten zwischen ambulanten und stationären Behandlungen, aber auch zwischen ambulanten Psychotherapien unterschiedlicher Therapieschulen, sind die Zeiträume zur Definition der Early Response nur schwer vergleichbar (vgl. Abschnitt 2.5.4).

Bei der Betrachtung der wöchentlichen Verbesserungsraten fällt auf, dass generell ein relativ großer Patientenanteil in der frühen Phase der Behandlung (etwa bis zur dritten Woche) eine relevante Verbesserung erzielt. Schon nach der ersten Woche erreicht ein Anteil von knapp einem Drittel der Patienten eine Verbesserung mit mittlerer Effektstärke. Selbst das „härteste“ Kriterium der klinisch signifikanten Verbesserung, entsprechend der konservativsten der hier verwendeten Definitionen der Response, wird nach der dritten Woche von 34% der Patienten mindestens einmal erreicht. Es ist fraglich, welche Aussagekraft diese hier erfassten punktuelle Verbesserung gegenüber einer anhaltenden Verbesserung haben. Gerade im stationären Setting sind Effekte, die am Anfang der Behandlung auftreten, und nicht direkt auf die Behandlung zurück gehen, denkbar: Durch die lokale Veränderung und die Trennung von ihrem Alltagsumfeld kann bei den Patienten ein gewisser Erholungseffekt, der zu einer kurzfristigen Verbesserung der Symptomatik führen könnte, auftreten. Zudem treffen die Patienten hier, anders als in der ambulanten Psychotherapie, zum Teil erstmals auf andere Patienten mit ähnlichen Erkrankungen. Dies könnte die akute Symptomentlastung nach dem Konzept der Wirksamkeit des sogenannten *therapeutischen Milieus* (Ammon, 1998) erklären: Der Schutzraum des stationären Settings, die starke Strukturierung des Alltags und die sozialen Beziehungen zwischen den Patienten können eine gewisse Entlastung hervorrufen, die folglich nicht explizit auf therapeutische Maßnahmen zurück geführt werden. Es ist jedoch fraglich, ob diese unspezifischen Milieueffekte mittel- bis langfristig anhalten.

Zusammenfassend erlauben die explorativen Verlaufgruppen eine bessere Vorhersage des längerfristigen Behandlungsergebnisses als die theoriegeleiteten Verlaufgruppen. Zur Bildung der explorativen Verlaufgruppen wurde die initiale Belastung und der gesamte Verlauf einbezogen, was zu einer besseren Differenzierung des Outcomes führt. Aufgrund der explorativen Vorgehensweise und der erstmaligen Untersuchung der Symptomverläufe ausschließlich im stationären Setting, ist die Replikation der Ergebnisse notwendig, um Implikationen aus diesen ableiten zu können. Die statistische Methode der Wachstumskurvenmodellierung, die zur explorativen Verlaufgruppenbildung angewendet wurde, wird im Abschnitt 6.3.3 diskutiert.

Für die theoriegeleiteten Verlaufsgruppen könnte sowohl der Zeitpunkt, für den eine Response a priori festgelegt wurde, als auch das inhaltliche Kriterium, nach dem eine relevante Verbesserung definiert wurde, nicht optimal für das untersuchte Setting und/oder die untersuchte Patientengruppe gewählt worden sein. Da in der Literatur ebenfalls keine einheitliche Definition der Early Response zu finden ist, und bisherige Ergebnisse sich hauptsächlich auf ambulante Behandlungen beziehen (Lambert, 2005), lassen sich die Ergebnisse der theoriegeleiteten Verlaufsgruppen schwer einordnen. Die Adäquatheit des Kriteriums, mit der eine relevante Verbesserung definiert wird, hängt gewiss auch von der Angemessenheit des untersuchten Konstrukts und des dafür eingesetzten Instruments ab. Die Relevanz der Verlaufs- und Ergebnismaße wird in der methodischen Diskussion in Abschnitt 6.3.6 diskutiert.

6.3 Methodische Diskussion

6.3.1 Alpha-Fehler und Teststärke

In der vorliegenden Untersuchung wurde ein 95%iges Signifikanzniveau gewählt, Ergebnisse mit einer 10% Alphafehlerwahrscheinlichkeit wurden ebenso wie die 1%ige und 0.1%ige Fehlerwahrscheinlichkeiten dargestellt. Auf eine Alpha-Adjustierung wurde dabei verzichtet. Begründet war dieses Vorgehen darin, dass bei dem weitgehend explorativen Charakter der ersten Untersuchungsfragestellung, auf deren Ergebnisse Fragestellung 2 und 3 aufbauen, eine möglichst hohe Teststärke angestrebt wurde. Damit wurde die Möglichkeit einer Alphafehler-Inflation in Kauf genommen, um auch geringe Effekte entdecken zu können. Das bedeutet, dass das Risiko, fälschlicherweise Zusammenhänge aufzudecken, die in der Population gar nicht vorhanden sind, dem Risiko, fälschlicherweise in der Population vorhandene Zusammenhänge zu übersehen, vorgezogen wird (Minimierung des Beta-Fehlers).

Dieses Vorgehen wurde insofern als angemessen eingeschätzt, da Symptomverläufen in der stationären Behandlung bislang wenig untersucht wurden und infolgedessen eine stärkere Fokussierung auf dem Erkenntnisgewinn als auf der Absicherung der gefundenen Zusammenhänge lag. Demzufolge besteht jedoch die Möglichkeit, dass ein Teil der gefundenen Zusammenhänge auf den Zufall zurückzuführen ist. Entsprechend wurden vor allem die Ergebnisse, die nicht konform mit der theoretischen Fundierung gehen, keinesfalls als Widerlegung der Theorie verstanden. Vielmehr liefern diese Ergebnisse lediglich einen Hinweis darauf, diese Merkmale in folgenden Untersuchungen nicht unberücksichtigt zu lassen. Dies betrifft vor allem die in Fragestellung 2 und 3 mehrfach durchgeführten Regressionsanalysen. Hinsichtlich der Effekte über die Gesamtmodelle, auf die sich die Hypothesen ausschließlich bezogen haben, besteht ein geringes Risiko der Alphainflation.

Betrachtet man jedoch die Effekte einzelner Prädiktoren, ist vor allem bei den nur tendenziell signifikanten Prädiktoren Vorsicht geboten. Aus diesem Grund wurden nur die Zusammenhänge oder Prädiktoren inhaltlich interpretiert, die das 95% Signifikanzniveau oder eine mindestens kleine Effektstärke erreichten.

6.3.2 Studiendesign, interne und externe Validität

Die vorliegende Studie wurde als naturalistische Studie im *Eingruppen-Prä-Post-Design* durchgeführt und ist den Effectiveness-Studien zuzuordnen. Die Vor- und Nachteile von Effectiveness-Studien wurden bereits im Theorieteil erläutert (vgl. Abschnitt 2.2). An dieser Stelle wird das Studiendesign und folglich die interne und externe Validität der vorliegenden Untersuchung kritisch betrachtet.

Ziel der Studie war es, möglichst präzise Aussagen über den Zusammenhang von Prozess und Wirksamkeit der Behandlung in der realen Versorgungssituation abzuleiten, die möglichst eindeutig auf den untersuchten Gegenstand, den Symptomverlauf während der Behandlung, zurückzuführen sind. In der vorliegenden Untersuchung ist es gelungen, in einem naturalistischen Studiendesign konsekutiv alle behandelten Patienten einzuschließen, die sich zur Studienteilnahme bereit erklärten. Zudem wurden Repräsentativitätsanalysen für die Untersuchungsstichproben durch Vergleiche zur Klinikpopulation durchgeführt, die in Abschnitt 6.1 beschrieben wurden. Dabei wiesen die teilnehmenden Patienten nur geringfügige Unterschiede zur Klinikpopulation auf. Um zudem die Übertragbarkeit der Ergebnisse auf das entsprechende stationäre Gesamtsetting der Rehabilitation psychischer/psychosomatischer Erkrankungen und der psychosomatisch/psychotherapeutischen Krankenhausversorgung zu erhöhen, wurden die teilnehmenden Kliniken nach Kriterien der Übertragbarkeit auf diese beiden Settings ausgewählt. Insgesamt kann aufgrund des naturalistischen Studiendesigns, der konsekutiven Stichprobe, der Absicherung der Selektion hinsichtlich der teilnehmenden Patienten und der Auswahl der Kliniken hinsichtlich des Settings von einer hohen *externen Validität* ausgegangen werden. Die Ergebnisse sind weitgehend auf einen großen Teil der stationären Versorgung von Patienten mit psychischen Erkrankungen (exklusive der psychiatrisch/psychotherapeutischen Krankenhausversorgung) übertragbar. Einschränkungen der Repräsentativität zur untersuchten Klinikpopulation beziehen sich auf die Behandlungsdauer und die Diagnosegruppen. Diese hängen mit dem gewählten Einschlusskriterium der begrenzten Behandlungsdauer zusammen, die gewählt wurden, welches wiederum zum Zweck einer erhöhten Repräsentativität für den gesamten stationären psychosomatisch/psychotherapeutischen Versorgungsbereich (Rehabilitation und Krankenhausbehandlung) angewendet wurde (vgl. Abschnitt 4.2.3 und 6.1.1).

Eine hohe *interne Validität*, also die Minimierung des Einflusses möglicher Störfaktoren, die die Ergebnisse konfundieren könnten, wurde durch die statistische Berücksichtigung von Confoundervariablen erreicht. Dazu wurde in der Hauptfragestellung (Vorhersage des Outcomes durch die Symptomverlaufgruppen) eine Vielzahl an empirisch oder inhaltlich begründeten Merkmalen in die Analysen eingeschlossen. Aufgrund des explorativen Charakters der ersten Fragestellung sollten möglichst viele inhaltlich hergeleitete Merkmale als Prädiktoren, bei Berücksichtigung einer ausreichenden statistischen Power, untersucht werden. Die interne Validität wurde demnach durch die statistische Berücksichtigung von Merkmalen, die in einem potentiellen Zusammenhang zum Behandlungsergebnis stehen, die jedoch nicht den Untersuchungsgegenstand darstellen, erhöht. Problematisch stellt sich hierbei die Auswahl der Untersuchungsmerkmale dar. Wie in Abschnitten 2.4 und 2.5 beschrieben, wird eine Vielzahl von Merkmalen diskutiert, die potentiell mit dem psychotherapeutischen Behandlungsergebnis zusammenhängt. Neben Patienten- und Prozessmerkmalen kommen auch Therapeutenmerkmale oder Merkmale der Behandlung bzw. Institution sowie interaktionsbezogene Merkmale in Betracht. In der vorliegenden Untersuchung sollten möglichst umfassend und vollständig die Merkmale als Prädiktoren untersucht werden, für die empirische Evidenz oder eine inhaltlich Rechtfertigung besteht. Neben einer Vielzahl von Patientenmerkmalen wurden die vier Prozessmerkmale Behandlungsdauer, Dosis und Art der psychotherapeutischen Maßnahmen sowie die Güte der therapeutischen Beziehung einbezogen. Dennoch kann nicht ausgeschlossen werden, dass die gefundenen Effektstärken des Zusammenhangs zum Outcome auch über andere, hier nicht untersuchte Merkmale vermittelt wurden. Um mögliche institutionsbedingte Effekte zu kontrollieren, wurde die behandelnde Klinik als Kontrollvariable eingeschlossen. Andere Strukturmerkmale wurden jedoch nicht berücksichtigt, da deren Erhebung einen erhöhten Aufwand zur Folge gehabt hätte, der durch die fehlende empirische Grundlage nicht gerechtfertigt schien. Die Auswahl der untersuchten Merkmale, die als Confounder in die Analysen eingeschlossen werden, muss auch unter statistischen Gesichtspunkten erfolgen: Der Einschluss einer Vielzahl von Prädiktoren geht auf Kosten der statistischen Teststärke, weshalb die Anzahl der Prädiktoren begrenzt sein sollte (vgl. Barth, 1997; sowie Abschnitt 4.6).

Einen weiteren Diskussionspunkt, der durch das Studiendesign bedingt wird, stellt die Besonderheit der Veränderungsmessung im Eingruppen-Prä-Post-Design dar: Durch das Fehlen einer unbehandelten Kontrollgruppe können die durch eine Intervention erzielten Veränderungen in diesem Design nicht kausal interpretiert, also nicht ausschließlich auf die Behandlung selbst zurückgeführt werden. Zu den weiteren möglichen Ursachen der Veränderungen zählen ein natürlicher Krankheitsverlauf im Sinne des von Eysenck postulierten Phänomens der Spontanremission (Eysenck, 1952; McNeilly & Howard, 1991),

ebenso wie die statistische Regression zur Mitte (Zwingmann & Wirtz, 2005). Die Regression zur Mitte beschreibt die statistische Tendenz, dass Individuen, die bei einer ersten Messung Extremwerte in einem Merkmal aufweisen, bei Wiederholungsmessungen geringere Abweichungen vom Gruppenmittelwert erzielen. Dabei handelt es sich um eine artifizielle zufallsbedingte Veränderung, die zu einer Überschätzung der Behandlungseffekte führen können. Dies trifft vor allem auf die Verlaufsuntersuchung von Subgruppen zu, da Subgruppen mit extremen Ausgangswerten⁴² eine erhöhte Wahrscheinlichkeit der Regression zur Mitte aufweisen (Zwingmann & Wirtz, 2005). Dies muss in den vorliegenden Ergebnissen insbesondere bei den explorativen Verlaufsgruppen berücksichtigt werden, da diese unter anderem anhand der Ausgangswerte der Symptombelastung gebildet wurden. Dennoch wird davon ausgegangen, dass die Unterschiede zwischen den explorativen Verlaufsgruppen und der resultierende Anteil an erklärter Varianz im längerfristigen Outcome aufgrund der großen Effektstärken auch unter Berücksichtigung möglicher statistischer Artefakte noch als relevant zu bewerten sind.

6.3.3 Kritische Auseinandersetzung mit der Methode des Growth Mixture Modeling

Das Ziel der explorativen Verlaufsanalyse war die Identifikation von distinkten Gruppen, die die individuelle Variabilität über die Zeit zu homogenen Subgruppen zusammenfassen. Bei der Growth Mixture Modeling Methode handelt es sich um ein exploratives Verfahren. Kritisch anzumerken ist, dass die Verlaufsgruppen demnach rein empirisch bestimmt wurden. Allerdings besteht derzeit nur eine sehr geringe theoretische Basis zu interindividuellen Unterschieden der Veränderungen während der psychotherapeutischen Behandlung. Ein Vorteil der GMM-Methode im Vergleich zu Endpunktmessungen liegt darin, dass letztere eben nur einen Zeitpunkt einbeziehen, der mit höherer Wahrscheinlichkeit von anderen Faktoren beeinflusst wird als von der Behandlung allein (vgl. auch Abschnitt 6.3.2). Durch die Hinzunahme mehrerer Messzeitpunkte erhöht sich die Validität der Ergebnismessung (Uher et al., 2010), was besonders in Beobachtungsstudien einen großen Vorteil darstellt. Ferner besteht bei den Verlaufsanalysen eine geringere Wahrscheinlichkeit der Verzerrung durch fehlende Werte, wie sie bei Endpunktmessungen auftreten können (vgl. auch Abschnitt 4.5.2).

Die Untersuchung des gesamten Verlaufs der Veränderung der Symptombelastung könnte ausschlaggebend sein für die Erkenntnis, wie und für welche Patienten die stationäre psychotherapeutische Behandlung wirkt bzw. nicht wirkt. Das in anderen Studien und mit anderen Methoden gefundene Phänomen der Early Response konnte auch mittels der explorativen Analyse gefunden werden. Die hohe Übereinstimmung der explorativ

⁴² Die Extremwerte beziehen sich auf die relative Lage zu dem Gesamtgruppenmittelwert.

identifizierten Early-Response_{EG}-Gruppe mit dem Erreichen des a priori festgelegten Kriteriums einer Early Response_{TG} festigt die Aussagekraft der explorativen Verlaufgruppen. Eindrücklich konnte in den Verlaufs-Outcome-Analysen gezeigt werden, welchen Vorhersagewert die mittels GMM identifizierten Verlaufgruppen für das katamnestische Outcome aufweisen, verglichen mit üblicheren Erfolgskriterien, an die die theoriegeleiteten Verlaufgruppen angelehnt waren. Mit GMM konnten drei Patientensubgruppen identifiziert werden, die alle mit einer ähnlich hohen Ausgangsbelastung in die Klinik kommen, dann aber im Verlauf der Behandlung gravierende Unterschiede der Veränderung aufweisen. Katamnestisch bestehen für das primäre Outcome Symptombelastung Unterschiede im Ausmaß einer großen Effektstärke und für das sekundäre Outcome gesundheitsbezogene Lebensqualität im Ausmaß einer mittleren Effektstärke zwischen den Gruppen. Für einen Teil der behandelten Patienten können mit dieser Methode sehr differenzierte Aussagen zum Verlauf und Outcome gemacht werden, die erste wertvolle Hinweise für die Optimierung von Behandlungsprozessen liefern (s. Abschnitt 6.4).

Einschränkend sind, wie schon in Abschnitt 4.5.2 aufgeführt, die derzeit verfügbaren Methoden zur Überprüfung des Modellfits der GMM-Analysen zu nennen: Die in der vorliegenden Arbeit angewendeten Verfahren zur Prüfung der Modellgüte und zur Entscheidung über die beste Klassenlösung werden zwar als robust und zuverlässig diskutiert, liefern aber keine abschließende und unumstrittene Antwort auf die Frage der Beurteilung der optimalen Anzahl an latenten Klassen (Jung & Wickrama, 2008; Stulz et al., 2010). Zudem wird empfohlen, die Ergebnisse von GMM-Analysen immer vor einem theoretischen Hintergrund und nicht ausschließlich anhand statistischer Parameter zu interpretieren (Muthén, 2004). In Abschnitt 5.2.3 wurden die identifizierten Klassen zwar nach inhaltlichen Kriterien modifiziert, allerdings liegen bislang keine fundierten klinisch-theoretischen Konzeptionen zu interindividuellen Unterschieden des Verlaufs der Belastung während der psychotherapeutischen Behandlung vor (Lutz et al., 2007b). Demzufolge können die Ergebnisse der GMM-Analysen nur schwer an theoretischen Modellen geprüft werden. Ein Versuch der theoretisch-konzeptuellen Annäherung wurde in der vorliegenden Arbeit mit der Bildung der theoriegeleiteten Verlaufgruppen und dem Vergleich beider Gruppenbildungen vorgenommen. Allerdings zeigt die bereits diskutierte Uneinheitlichkeit der Definition und Operationalisierung von Verlaufgruppen (vgl. Abschnitt 6.2.1 und 6.2.3.4) und die methodische Unterschiedlichkeit der beiden Verfahren der Gruppenbildung (theoriegeleitet und explorativ) die Grenzen der Eignung als theoretische Grundlage auf. Eine Erweiterung der theoretischen Konzeptionalisierung von interindividuellen Behandlungsverläufen, unter Umständen auch gestützt auf bisherige empirische Modelle, könnte die Forschung auf diesem Gebiet bereichern.

Ferner muss kritisch angemerkt werden, dass für den größten Teil der untersuchten Stichprobe, die Linear-Response_{EG}-Gruppe, keine Differenzierung der Symptomverläufe vorgenommen werden konnte. Neben der Folgerung, dass ein großer Teil der stationär behandelten Patienten keine Differenzierung im Verlauf der Symptombelastung aufweist, können auch methodische Erklärungsansätze erwogen werden. In den GMM-Analysen wurden die Parameter Intercept, Slope und quadratischer Slope einbezogen. Denkbar wäre auch die Untersuchung kubischer Funktionen, mit denen beispielsweise ein Verlauf mit einer Verbesserung in der Anfangsphase, gefolgt von einem Wiederanstieg der Symptomatik im weiteren Verlauf und einem erneuten Rückgang der Symptombelastung zum Ende der Behandlung abgebildet werden könnten. Solche oder ähnliche Verlaufsmuster konnten mit dem hier berechneten Modell nicht überprüft werden. Allerdings fehlen auch hier die theoretischen Annahmen hinter solchen Verläufen, weshalb auf den Einschluss zusätzlicher Parameter verzichtet wurde. Eine weitere mögliche Differenzierung von Verläufen wäre über die Häufigkeit und Art der wöchentlich auftretenden Schwankungen der Symptombelastung, angelehnt an das Konzept der Sudden Gains (Tang & DeRubeis, 1999b; Tang et al., 2007), denkbar (vgl. auch Abschnitt 2.5.5). Bei einer ausreichend großen Stichprobe könnte alternativ vorab eine Einteilung des initialen Schweregrades, wie es Lutz et al. (2007b) vorschlagen, vorgenommen werden, um anschließend diese a priori definierten Patientengruppen in Hinblick auf homogene Subgruppen (hinsichtlich des Verlaufs) zu untersuchen. Diese Vorgehensweise könnte in weiterführenden Untersuchungen zur Replikation der hier identifizierten Verlaufsgruppen angedacht werden: Trotz der relativ großen Stichprobe, die hier untersucht werden konnte, fallen drei Gruppen relativ klein aus (5% - 11%), während sich die vierte und mit Abstand größte Gruppe der Linear Response (71%) von den anderen drei Gruppen stark durch die initiale Symptombelastung unterscheidet. Durch die Vorabenteilung in Gruppen des initialen Schweregrades könnte ein höherer Informationsgewinn erzielt werden.

Insgesamt erweist sich das Growth Mixture Modeling trotz der genannten Einschränkungen als derzeit adäquate Methode, um homogene Subgruppen der Verläufe abzubilden. Vorteile der Methode stellen die Annahme distinkter Verlaufsgruppen innerhalb einer Population, die Flexibilität und Erweiterbarkeit (z.B. hinsichtlich nichtlinearer Verläufe) und die Robustheit gegenüber fehlenden Werten dar (Keller & Hautzinger, 2007; Lutz et al., 2009; Lutz et al., 2007b; Stulz et al., 2010).

6.3.4 Validität der Prädiktormasse

6.3.4.1 *Validität der Diagnosen*

Insgesamt zeigen die Ergebnisse der zweiten und dritten Fragestellung, dass die Diagnose der Patienten zum Teil zwar mit dem Behandlungsverlauf, nicht jedoch mit dem katamnestischen Behandlungsergebnis zusammenhängt. Eine Ausnahme bildet die Anpassungsstörung, die mit einer geringen Symptombelastung zur Katamnese einhergeht. In indikationsvergleichenden Auswertungen wurden häufig sich deutlich unterscheidende Effektstärken für verschiedene Diagnosegruppen gefunden (Steffanowski, 2007; Barghaan, 2005 vgl. auch 2.4.1). In der vorliegenden Studie besteht nur ein schwacher Zusammenhang zwischen der Diagnose und dem Behandlungsergebnis. Eine Erklärung für den geringen Zusammenhang zwischen Diagnose und Symptombelastung (im Verlauf und als Outcome) liegt in der Abgrenzung und der Validität der Diagnosestellung: In dieser Studie wurden für die Einteilung in Diagnosegruppen für jeden Patienten alle vergebenen Diagnosen herangezogen. Alle Patienten, für die als Zweit- oder Drittdiagnose (bis zu fünf Diagnosestellen) eine depressive Erkrankung kodiert wurde, wurden ebenso in die Gruppe der depressiven Erkrankungen eingeschlossen wie die Patienten, die eine Depression als Erstdiagnose hatten. Dies führt zu einer größeren Vermischung der Störungsbilder, die klare Abgrenzung der Diagnosegruppen für Vergleiche unterschiedlicher Störungsgruppen ist erschwert. Diese Vorgehensweise kann kritisch diskutiert werden: Die Entscheidung, nicht ausschließlich die Erstdiagnose zu betrachten, war zum einen dadurch begründet, dass die Behandlung sich laut Leitlinien ebenfalls nicht ausschließlich auf die Erstdiagnose bezieht bzw. beziehen sollte (Deutsche Rentenversicherung Bund, 2011). Gerade Patienten mit psychischer Komorbidität bedürfen einer darauf abgestimmten Behandlung. Zum anderen erscheint es generell fraglich, ob die Rangfolge der vergebenen Diagnosen (Erst- vs. Zweitdiagnose) bei psychischen Erkrankungen eine adäquate Beschreibung der Psychopathologie darstellt. Die Frage, welche Erkrankung bei Komorbidität als primär bezeichnet wird und entsprechend zur Behandlung indiziert, stellt sich vor allem bei psychischen Erkrankungen als schwierig dar. Zudem weist die Häufigkeit des Auftretens von komorbiden psychischen Störungen auf eine erschwerte Abgrenzung zwischen unterschiedlichen Störungsbildern hin (Baumeister, Kury & Bengel, 2011).

Ferner müssen die Ergebnisse zu den Diagnosegruppen unter der Einschränkung der Validität der Diagnosestellung interpretiert werden. Die Diagnosen wurden aus den Entlassberichten der Kliniken entnommen, um eine klinisch möglichst fundierte und valide Einschätzung zu erhalten. Der Entlassdiagnose beruht, anders als die Aufnahmediagnose, auf der Einschätzung von Psychotherapeuten und Fachärzten, die den Patienten über mehrere Wochen behandelt haben. Dennoch stützt sich die Diagnostik in den Kliniken auf

ein klinisches Urteil, welches nicht, wie beispielsweise in Studienstichproben, durch ein psychometrisch überprüftes, standardisiertes klinisches Interview abgesichert wird. Die Ergebnisse der vorliegenden Arbeit sind entsprechend nur bedingt vergleichbar mit Ergebnissen störungsspezifischer Studien, die zum einen die Beschreibung der Störung ausschließlich auf die Primärdiagnose beziehen und zum anderen eine Absicherung der Diagnose vorgenommen haben. Schließlich beziehen sich unterschiedlich hohe Effektstärken verschiedener Indikationsgruppen meist auf störungsspezifische Maße. Die Effektstärken der Veränderung in generischen Outcome-Maßen unterscheiden sich zwischen den Diagnosegruppen nicht in dem Ausmaß wie die Effektstärken störungsspezifischer Instrumente (Steffanowski et al., 2007).(Schulz et al., 2008)

6.3.4.2 *Validität der Erfassung der Behandlungsdosis*

Die Daten zur Dosis und Art der Behandlungsmaßnahmen basierten auf den Kodierungen der Klassifikation therapeutischer Leistungen (Deutsche Rentenversicherung Bund, 2007). Mithilfe des Klassifikationssystems wird eine stark differenzierte Kategorisierung von Behandlungsmaßnahmen ermöglicht. Dennoch bleibt unklar, inwiefern die KTL-Kodierungen ein valides Abbild der Behandlungen in den Kliniken darstellen. Die Validität des Instruments ist bislang noch nicht überprüft. Aufgrund dieser Einschränkung wurden die Dosisanalysen ausschließlich auf Basis einer groben übergeordneten Einteilung in psychotherapienahe und psychotherapieferne Behandlungsmaßnahmen durchgeführt. Da die Dosisanalysen nicht auf der Ebene von einzelnen Behandlungsmaßnahmen vorgenommen wurden, wird die Einschränkung der Validität der KTL als vernachlässigbar bewertet. Allerdings kann auch die Einteilung in psychotherapienahe und psychotherapieferne Maßnahmen, unabhängig von der Validität der KTL, kritisch diskutiert werden. Die Behandlung in den Kliniken umfasste ein breites Spektrum therapeutischer Maßnahmen, das neben Einzel- und Gruppenpsychotherapeutischen Behandlungen unter anderem Entspannungs-, Bewegungs-, Ergo- und Künstlerische Therapien einschließt (vgl. Abschnitt 2.1.2). Eine Einteilung in psychotherapienahe Behandlungen ist folglich nicht immer eindeutig vorzunehmen. So können beispielsweise Entspannungsverfahren sowohl zu den psychotherapienahen als auch zu den psychotherapiefernen Behandlungen gezählt werden. Um eine möglichst fundierte Einteilung zu erhalten, wurde diese an Kategorien angelehnt, die im Zuge der Leitlinienerstellung für die rehabilitative Behandlung von depressiven Patienten entwickelt wurden (Barghaan et al., 2010; Dirmaier et al., 2010).

Eine weitere Einschränkung der Angaben zur erhaltenen Behandlung ist dadurch bedingt, dass die medikamentöse Behandlung der untersuchten Patienten nicht berücksichtigt werden konnte. Es ist davon auszugehen, dass ein Anteil der Patienten im untersuchten Setting zusätzlich zu dem beschriebenen Behandlungsspektrum auch

psychopharmakologisch behandelt wurde. Systematische Effekte der Medikation auf die Verläufe ebenso wie auf das Outcome können nicht ausgeschlossen werden. Folgende Untersuchungen sollten die medikamentöse Behandlung differenziert erfassen und in die Prädiktoranalysen einbeziehen.

6.3.4.3 *Validität der therapeutischen Beziehung im stationären Setting*

Auf die Problematik der Einzeitpunktmessung zur Erhebung der Güte der therapeutischen Beziehung und die damit verbundene Frage nach der Wahl des Erhebungszeitpunktes wurde schon in Abschnitt 6.2.2.1 eingegangen. Eine weitere Einschränkung der Validität der eingeschätzten Güte der therapeutischen Beziehung ergibt sich aus dem Behandlungssetting, in dem die Studie durchgeführt wurde. Angesichts der Tatsache dass die Erhebung in ambulanten Psychotherapiestudien meist direkt im Anschluss an die Therapiesitzung stattfindet, kann entsprechend eine vergleichsweise unverzerrte Einschätzung gemessen werden. Im Rahmen dieser Studie konnte die Verlaufserhebung nicht an eine bestimmte therapeutische Maßnahme gekoppelt werden. Dies hatte zur Folge, dass die Zeitspanne seit der letzten psychotherapeutischen Sitzung zwischen den Patienten aber auch für einen Patienten zwischen den Wochen variieren konnte. Folglich könnten Verzerrungen durch Erinnerungseffekte oder durch kognitive Prozesse, wie etwa kognitive Anpassungsprozesse bei Dissonanzen zwischen therapeutischem Geschehen und dem eigenen Befinden, aufgetreten sein. Ähnlich wie bei der direkten Veränderungsmessung (vgl. Abschnitt 2.3.1) könnte die verzögerte Einschätzung der therapeutischen Beziehung durch das gegenwärtige Befinden und implizite Theorien zur Erklärung dessen verfälscht werden (Farin et al., 2009). Zudem ist die Validität dahingehend eingeschränkt, dass nicht klar war, auf welche psychotherapeutische Behandlung (z.B. Gruppen- oder Einzelpsychotherapie) bzw. auf welchen Psychotherapeuten sich die Einschätzung bezog. In dem Fragebogen wurde darauf hingewiesen, die Beurteilung der therapeutischen Beziehung für die psychotherapeutische Einzeltherapie bzw. den Bezugstherapeuten vorzunehmen. Aufgrund der Vielfalt der therapeutischen Angebote im stationären Rahmen können hier jedoch Unschärfen der Einschätzung der Güte der therapeutischen Beziehung aufgetreten sein.

6.3.5 *Angemessenheit der Erhebungszeitpunkte*

Kritisch zu betrachten ist die Validität der Baseline-Erhebung hinsichtlich der Wahl des ersten Messzeitpunktes, der hier zu Behandlungsbeginn, also kurz nach der Aufnahme in der Klinik, gewählt wurde. Durch Erwartungseffekte und die möglicherweise kurzfristig auftretende Veränderung der Symptombelastung (im Zuge der neuen Situation des bevorstehenden mehrwöchigen Klinikaufenthaltes), könnte es zu einer Verzerrung der Baseline-Erhebung kommen. Dies würde vor allem auf zeitinstabile Merkmale, die zustands-

und situationsabhängigen Veränderungen unterliegen (im Sinne eines State Merkmals), zutreffen. Nun ist aber für die vorliegende Fragestellung nach (Verbesserungs-)Verläufen gerade ein solches (krankheitsbezogenes) Merkmal notwendig, von dem eine hohe Änderungssensitivität über eine relativ kurze Zeit angenommen werden kann, wie es die Symptombelastung in diesen Indikationsbereichen darstellt (vgl. Abschnitt 2.3.1 und Abschnitt 4.3.1). Um die Validität der Baseline-Erhebung zu erhöhen, könnte ein Messzeitpunkt *vor* Behandlungsbeginn, oder im Idealfall schon vor Bewilligung der stationären Behandlung, angesetzt werden. Somit könnte der Bias durch zustandsbezogene Befindlichkeitsänderung bei Beginn der stationären Behandlung (beispielsweise motivationale und erwartungsbezogene Veränderungen, spontane Verbesserungen im Sinne eines „Erholungseffektes“) reduziert werden. Allerdings wären für eine solche Erhebung mehr Ressourcen und ein höherer organisatorischer Aufwand nötig, sodass der Gewinn an Validität und die Ökonomie gegeneinander abgewägt werden sollten.

Hinsichtlich der Verlaufgruppenbildung kann die hier vorgenommene Festlegung der Erhebungszeitpunkte ebenfalls kritisch betrachtet werden. Der explorativen Verlaufgruppenbildung wurden wöchentliche Erhebungen zugrunde gelegt. Die Veränderung der Symptombelastung wurde demnach an einem Zeitintervall und nicht an einem Dosisintervall gemessen. In den Untersuchungen zum Verlauf der ambulanten Behandlung kann nach jeder Psychotherapiesitzung eine Erhebung stattfinden. Da die Therapiesitzungen in einem festgelegten zeitlichen Intervall stattfinden, decken sich Zeit- und Dosisintervall hier meistens. Allerdings zeigte auch hier die Diskussion über Erklärungen für den in der Übersichtsarbeit von Illardi und Craighead (1994) gefundenen hohen Anteil an Early Response in der CBT, wie wichtig es ist, neben der Dauer auch die Dichte der Psychotherapie einzubeziehen, vor allem wenn diese im Laufe der Behandlung variiert (Tang & DeRubeis, 1999a; vgl. auch Abschnitt 2.5.5). Für die stationäre psychotherapeutische Behandlung wird zwar von einem proportionalen Verhältnis von Zeit und Dosis ausgegangen, die gefundene Korrelation zwischen Dauer und Dosis von $r = .38$ zeigt jedoch, dass dies nicht vollständig gegeben ist. Wie schon in Abschnitt 6.2.3.3 diskutiert, können die Implikationen aus den Ergebnissen eher auf die Behandlung als Ganzes denn auf einzelne Behandlungsmaßnahmen bezogen werden.

Das Outcome wurde katamnestisch 6 Monate nach der stationären Behandlung erhoben. Alternativ hätte auch das Ergebnis zur Entlassung als Outcome-Kriterium genommen werden können. Da aber der Wert der Symptombelastung zur Entlassung in die Bildung der Verlaufgruppen einbezogen wurde, wäre eine Vorhersage der Symptombelastung zur Entlassung tautologisch und wenig aufschlussreich. Ferner wird gerade bei der Behandlung von psychisch erkrankten Patienten die katamnestische Erhebung des Behandlungserfolgs

empfohlen, da nur so die Nachhaltigkeit und Verstetigung der in der Therapie erreichten Veränderung überprüft werden kann (Kendall et al., 2004).

6.3.6 Angemessenheit der Verlaufs- und Outcome-Maße

Sowohl für die Verlaufsmessung als auch für die Ergebnismaße stellt sich die Frage, ob die hier angewendeten Zielbereiche und Instrumente angemessen sind. Die patientenseitige Einschätzung der Symptombelastung ist ein weit verbreitetes Ergebnismaß in der stationären Psychotherapie (vgl. Abschnitt 2.3.1). Dennoch muss berücksichtigt werden, dass mit diesem Outcome-Maß nur ein Ausschnitt des subjektiven Befindens dargestellt werden kann. Gerade unter dem Gesichtspunkt der divergierenden Therapieziele bei unterschiedlichen therapeutischen Schulen können ebenso andere oder umfassendere Festlegungen des Therapieergebnisses, etwa die interaktionelle Beziehungsgestaltung, erwogen werden. Vorstellbar wären auch Zielgrößen wie z.B. die Veränderung des Krankheitskonzeptes oder die Erhöhung der Selbstwirksamkeitserwartung hinsichtlich psychischer Problembereiche. Zur Fundierung der Aussagekraft der Ergebnisse wurde das sekundäre Outcome, die psychischen Aspekte der gesundheitsbezogenen Lebensqualität, untersucht. Die unterschiedlichen Ergebnisse im primären und sekundären Outcome verdeutlichen, dass die erreichten Effektstärken ebenso wie die Vorhersage der Wirksamkeit über den Symptomverlauf von dem Zielkriterium abhängen. Der Vorteil der angewendeten Ergebnismaße ist deren weite Verbreitung (vgl. Abschnitt 2.3.1), was eine erhöhte Vergleichbarkeit der Ergebnisse mit denen anderer Studien ermöglicht.

Ferner kann auch die Verwendung der Symptombelastung als Verlaufmaß kritisch betrachtet werden. Für eine hochfrequente Messung, wie sie in der vorliegenden Studie durchgeführt wurde, ist ein Maß bzw. Instrument erforderlich, von dem eine hohe Veränderungssensitivität für diesen kurzen Zeitraum anzunehmen ist. Für den HEALTH-49, der zur Messung der Symptombelastung eingesetzt wurde, existieren bislang keine empirischen Daten zur wöchentlichen Veränderungsmessung. Da sich der Fragebogen aber durch eine zufriedenstellende Prä-Post-Veränderungssensitivität auszeichnet, und zudem aufgrund der Kürze der Module praktikabel und ökonomisch für eine hochfrequente Erhebung ist, wird er als angemessen für die Verlaufserhebung eingeschätzt (Rabung et al., 2009). Ein weiterer Vorteil der Messung der Symptombelastung ist deren häufiger Einsatz in der Psychotherapieforschung und insbesondere in Verlaufsanalysen psychotherapeutischer Behandlung (Keller & Hautzinger, 2007; Lutz et al., 2007b; Stulz et al., 2007; Uher et al., 2010), was eine Einordnung und Vergleichbarkeit der Ergebnisse aus diesem neuen Forschungsansatz erst ermöglicht.

6.4 Ausblick

Aus den Ergebnissen der vorliegenden Arbeit ergeben sich einige wichtige Implikationen für die weitere Erforschung von Behandlungsprozessen in der stationären psychotherapeutischen Behandlung. Durch die Identifikation von Symptomverlaufgruppen konnte hier erstmals im stationären Setting eine Unterteilung in distinkte Symptomverlaufgruppen vorgenommen werden, die einen bedeutenden Zusammenhang mit dem längerfristigen Behandlungsergebnis aufweisen. Aufgrund der explorativen Vorgehensweise müssen die Ergebnisse in zukünftige Untersuchungen überprüft werden. Bislang fehlt es an klinisch-theoretischen Konzepten zur psychotherapeutischen Behandlung, die sich auf interindividuell unterschiedliche Behandlungsverläufe beziehen. Ansätze wie die Early Response oder die Sudden Gains beinhalten zwar die Annahme, dass bestimmte Patienten(-gruppen) individuelle Verläufe aufweisen, in beiden Konzepten wird aber nur auf einen bestimmten Ausschnitt des Behandlungsprozesses eingegangen. In der vorliegenden Arbeit wurde versucht, dem Anspruch einer zusätzlichen theoriebasierten Überprüfung des Symptomverlaufs durch die Bildung der theoriegeleiteten Verlaufgruppen gerecht zu werden. Allerdings konnte gezeigt werden, dass sich diese Kriterien, die ebenfalls nicht den gesamten Verlauf einbeziehen, wenig zur Einordnung der explorativen Ergebnisse eignen. Es wird deutlich, dass sowohl die weitere Erforschung des Therapieprozesses als auch die Erweiterung von theoretischen Konzepten, an denen empirische Modelle überprüft werden könnten, erforderlich ist. Die über explorative Methoden erzielten Ergebnisse könnten die konzeptionelle Beschreibung des Behandlungsprozesses bereichern, da mit ihnen die Darstellung komplexer Behandlungsprozesse möglich ist (Lutz et al., 2007b).

Hinsichtlich der untersuchten Prädiktor- und Outcome-Maße wurden die Einschränkungen der vorliegenden Studie und die Verbesserungsmöglichkeiten für nachfolgende Arbeiten bereits in den jeweiligen Abschnitten beschrieben. Diese beziehen sich vor allem darauf, eine reliable und valide Erfassung der Diagnosen sowie der Art und der Dosis der Behandlungsmaßnahmen einzusetzen. Ferner wäre eine Untersuchung der Beziehung zwischen Prozessmerkmalen, wie dem Verlauf der Symptombelastung und dem Verlauf der therapeutischen Beziehung denkbar, die ein besseres Verständnis therapeutischer Wirkmechanismen und -prozesse ermöglichen könnte.

Die Identifikation von distinkten Veränderungsmustern und die Möglichkeit, über diese Veränderungsverläufe das längerfristige Behandlungsergebnis vorherzusagen, liefert zudem eine Reihe möglicher Implikationen für die klinische Praxis: Zunächst ist anzuführen, dass die hochfrequente Untersuchung von Veränderungen im Verlauf der Behandlung einen Vorteil in der klinischen Praxis erbringen könnte, wenn die Messergebnisse im Sinne eines Monitorings direkt an den behandelnden Therapeuten zurückgemeldet werden und dieser

dementsprechend zeitnahe Modifikationen der Behandlung vornehmen könnte. Für die therapeutische Arbeit können die Kenntnisse über potentielle Verläufe zur frühzeitigen Adaptation der Behandlung eingesetzt werden. So könnten beispielsweise für die Patienten der hier gefundenen Delayed-Response-Gruppe ebenso wie für die Patienten, die während der gesamten Behandlung nicht ansprechen, frühzeitige Veränderungen schon während der Behandlung angedacht werden. Für die Delayed Response könnte verstärkt auf die Integration der während der Behandlung eingeleiteten Prozesse in den Alltag der Patienten eingegangen werden. Scheinbar profitieren diese Patienten, die hoch belastet in die Klinik kommen, erst zum Behandlungsende in einem substantiellen Ausmaß von der Behandlung. Die erreichte Entlastung kann aber nicht über den Zeitraum eines halben Jahres aufrechterhalten werden. Die hohe initiale Symptombelastung und das späte und instabile Ansprechen könnten auch eine Indikation für verstärkte Nachsorge sein. Für diese Patienten kann eine stärkere Fokussierung auf die Weiterführung der Behandlung angemessen sein. Denkbar wären Versorgungskonzepte, in denen der Patienten bei der Suche und der Auswahl ambulanter psychotherapeutischer Angebote sowohl auf der inhaltlichen Informationsebene als auch motivational unterstützt wird. Des Weiteren könnte auch die Entwicklung und Implementierung von gezielt niedrighschwelligem psychotherapeutischen Angeboten, mit denen mögliche Wartezeiten auf einem ambulanten Psychotherapieplatz überbrückt werden können, angedacht werden.

Nicht auszuschließen ist sicherlich auch die Überlegung, schon die stationäre Behandlung für diese Patienten auszudehnen, um langzeitstabile Behandlungseffekte zu gewährleisten. Die Annahme einer stabileren Response durch eine längere Behandlung konnte hier zwar nicht überprüft werden, die Ergebnisse aus den Studien zur Early Response können allerdings auch dahingehend interpretiert werden: Möglicherweise ist es nicht (alleine) das frühe Ansprechen, dass mit einem positiven längerfristigen Outcome zusammenhängt, sondern vielmehr die Möglichkeit, diese Veränderung noch während der Behandlung zu stabilisieren. In den meisten Studien, in denen Patienten(-gruppen) gefunden wurden, die früh ansprechen und gute Behandlungsergebnisse aufwiesen, wurden diese Patienten noch weiter behandelt. So kann auch für die hier gefundene Gruppe der Early Response nicht einfach geschlussfolgert werden, dass die erzielten Verbesserungen auch bei einer kürzeren Behandlung gefunden werden. Ginge man der Annahme nach, dass weniger das frühe Ansprechen und mehr die Stabilisierungsphase *nach* einer relevanten Verbesserung ausschlaggebend für ein gutes Therapieergebnis ist, so müsste man gerade für die Patienten mit einer Delayed Response längere Behandlungen erwägen. Die Patienten der Non-Response-Gruppe müssten entweder auch so lange behandelt werden, bis sie auf die Behandlung ansprechen und ausreichend stabilisiert sind. Ein anderer Ansatz wäre, diese Patienten möglichst früh in der Behandlung zu identifizieren und Modifikationen der

Behandlung vorzunehmen. Offensichtlich können diese Patienten von der Behandlung, wie sie sie erhalten haben, nicht ausreichend profitieren. Unter Umständen würden sie auch bei längeren Behandlungen keine Veränderung der Symptombelastung zeigen. Folglich könnte eine frühzeitige Veränderung der Behandlung angedacht werden.

Eine weitere Implikation stellt die Information der Patienten zum Zweck einer realistischeren Erwartungshaltung hinsichtlich der zeitlichen Wirksamkeit der Therapie dar: Der Therapeut könnte den Patienten mit einer potentiellen Delayed Response darüber informieren, dass eine Veränderung der Symptomatik erst nach einer gewissen Zeit eintritt und somit die Compliance des Patienten und die kontinuierliche Fortsetzung der Behandlung gefördert werden. Gleichzeitig kann bei einer potentiellen Early Response schon zu Beginn der Behandlung darauf hingewiesen werden, dass eine Fortsetzung der Behandlung sinnvoll ist, selbst wenn eine frühe Verbesserung der Symptombelastung auftritt. Wird der Patient angemessen über den potentiellen Verlauf der Symptombelastung aufgeklärt, kann die Zufriedenheit seitens des Patienten und damit die Compliance erhöht werden.

Aus den Prädiktoranalysen lassen sich ebenfalls Hinweise für die klinische Praxis ableiten. Es könnten vor allem für Patienten mit niedrigem sozioökonomischen Status Veränderungen in therapeutischen Prozessen angedacht werden: Ein niedriger Schulabschluss ebenso wie die Erwerbslosigkeit zeigten sich durchgehend sowohl als relevante Prädiktoren eines ungünstigeren Verlaufs der Symptombelastung während der Behandlung als auch des längerfristigen Behandlungsergebnisses, sowohl bezogen auf die Symptombelastung als auch die gesundheitsbezogene Lebensqualität. Mögliche Gründe für den Zusammenhang zwischen dem sozioökonomischen Status und dem Behandlungsergebnis wurden im letzten Abschnitt diskutiert. Die Annahme, dass Patienten mit einem niedrigeren sozioökonomischen Status einen schlechteren Zugang zur psychotherapeutischen Behandlung haben, konnte hier nicht bestätigt werden. Selbst unter Berücksichtigung der erhaltenen Dosis psychotherapienaher Maßnahmen blieb der Zusammenhang zwischen niedrigem Schulabschluss und Erwerbslosigkeit mit einem schlechteren Outcome bestehen. Dementsprechend sollten die psychotherapeutischen Verfahren möglicherweise hinsichtlich der Eignung und Passung für Patienten mit niedrigem sozioökonomischem Status überdacht werden. Für die klinische Praxis könnte eine besondere Aufmerksamkeit hinsichtlich der Behandlungsprozesse bei diesen Patienten und gegebenenfalls eine entsprechende Anpassung der Therapien an die Anforderungen und Bedarfe der Patienten erwogen werden.

Eine weitere relevante Gruppe an Prädiktoren des Symptomverlaufs und des Behandlungsergebnisses bilden die Merkmale, die auf chronifizierte Erkrankungen hinweisen. Dieses Ergebnis ist konsistent mit bisherigen Befunden zu Prädiktoren der Wirksamkeit stationärer psychotherapeutischer Behandlung. Chronifizierte Patienten weisen

häufig langfristigen Krankheitsverläufen und Behandlungswegen auf (Potreck-Rose & Koch, 1994). Es wird angenommen, dass der langfristige Verlauf der Erkrankung und möglicherweise auch inadäquate und wenig effektive Vorbehandlungen bei dieser Patientengruppe zur Folge haben könnten, dass sie schwerer zu behandeln ist. Folglich sollte hier ein Schwerpunkt in der verbesserten *frühzeitigen* Identifikation chronisch erkrankter Patienten liegen, um Chronifizierungsprozesse schon zu einem früheren Zeitpunkt aufzuhalten oder zumindest abschwächen zu können.

Eine übergeordnete Fragestellung der vorliegenden Arbeit stellte neben der Untersuchung der Symptomverläufe die Untersuchung des Zusammenhangs der Behandlungsdauer mit dem Behandlungsergebnis dar. Aus den Ergebnissen lassen sich nur wenige Ableitungen für die Versorgungspraxis formulieren. Wurden relevante Patienten- und Prozessmerkmale sowie die Symptomverlaufgruppen bei der Vorhersage des Behandlungsergebnisses berücksichtigt, stellte die Behandlungsdauer einen bedeutenden positiven Prädiktor des Behandlungserfolgs dar. Da dieser Zusammenhang nur in Kombination mit anderen Merkmalen auftritt, sollten weitere Indikatoren, wie die initiale Belastung, die Chronifizierung, Komorbiditäten und vor allem auch der Behandlungsprozess zur Bestimmung der Behandlungszeiten herangezogen werden. Zusammengefasst weisen die Ergebnisse auf die Notwendigkeit hin, die stationäre psychotherapeutische Behandlung und dementsprechend auch die Dauer der Behandlung unter Berücksichtigung von Therapieprozessen stärker zu individualisieren.



7 Literaturverzeichnis

- Altman, D.G., Trevor, B., Gardner, M. & Machin, D. (2000). *Statistics with Confidence: Confidence intervals and statistical guidelines (2 ed.)*. London: BMJ Books.
- Ammon, G. (1998). *Dynamische Psychiatrie (3 ed.)*. Eschborn: Klotz.
- Anderson, E.M. & Lambert, M.J. (2001): A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology, 57(7)*, 875-888.
- Andrade, L., Caraveo-Anduaga, J.J., Berglund, P., Bijl, R., Kessler, R.C., Demler, O., Walters, E., Kylyc, C., Offord, D., Ustun, T.B. & Wittchen, H.U. (2000): Cross-national comparisons of the prevalences and correlates of mental disorders. *Bulletin of the World Health Organization, 78(4)*, 413-426.
- Andreas, S., Dirmaier, J., Harfst, T., Kawski, S., Koch, U. & Schulz, H. (2009): Development and evaluation of a case group concept for inpatients with mental disorders in Germany: using self-report and expert-rated instruments. *European Psychiatry: the Journal of the Association of European Psychiatrists, 24(2)*, 105-111.
- Andreas, S., Dirmaier, J., Koch, U. & Schulz, H. (2003): DRG-Systeme in der Versorgung von Patienten mit psychischen Störungen: Zur Konzeption eines Klassifikationssystems für Fallgruppen. *Fortschritte der Neurologie-Psychiatrie, 5*, 234-243.
- Andreas, S., Dirmaier, J., Koch, U. & Schulz, H. (2005). *Fallgruppen in der stationären Rehabilitation von Patienten mit psychischen/psychosomatischen Erkrankungen. Unveröffentlichter Abschlussbericht des Projektes C2 im NVRF*. Hamburg: Institut und Poliklinik für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf.
- APA – American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV-TR (4th edition, Text Revision)* Washington, DC: American Psychiatric Association.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung (Vol. 10)*. Berlin: Springer.
- Barghaan, D., Harfst, T., Watzke, B., Dirmaier, J., Koch, U. & Schulz, H. (2007): Merkmale stationärer psychotherapeutischer Versorgung in Deutschland. *Psychotherapie im Dialog, 8(1)*, 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. *Gruppenpsychotherapie und Gruppendynamik, 45*, 83-103.
- Barghaan, D., Schulz, H., Steinmann, M., Krattenbacher, T., Koch, U. & Dirmaier, J. (2010). Entwicklung von Therapiestandards in der Rehabilitation von Patienten mit Depression - zwischen Evidenzbasierung und Praktikabilität. In F. Petermann (Ed.), *Neue Ansätze in der psychosomatischen Rehabilitation* (pp. 151-178). Regensburg: Roderer.

- Barghaan, D., Watzke, B., Koch, U. & Schulz, H. (2005). *Analyse von Ausgangs-, Prozess-, und Outcomedaten von Rehabilitationsmaßnahmen der BfA für Patienten mit psychischen/psychosomatischen Störungen, Unveröffentlichter Abschlussbericht*. Hamburg: Institut und Poliklinik für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf.
- Barkham, M., Rees, A., Stiles, W.B., Shapiro, D.A., Hardy, G.E. & Reynolds, S. (1996): Dose-effect relations in time-limited psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 64(5), 927-935.
- Barth, M. (1997). Konzeption und Evaluation multipler Regressionsanalysen in der anwendungsorientierten klinisch-psychologischen Forschung. In B. Strauß & J. Bengel (Eds.), *Forschungsmethoden in der Medizinischen Psychologie*. Göttingen: Hogrefe.
- Bassler, M., Krauthauser, H. & Hoffmann, S.O. (1995a): Welche Faktoren beeinflussen die Dauer von stationärer Psychotherapie? *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 45, 167-175.
- Bassler, M., Potratz, B. & Krauthauser, H. (1995b): Der Helping Alliance Questionnaire (HAQ) von Luborsky. *Der Psychotherapeut*, 40, 23-32.
- Baumeister, H. (2008): Anpassungsstörungen im Sinne einer reaktiven Depression - Charakteristika und Prävalenzraten der deutschen Allgemeinbevölkerung. *Psychosomatik und Konsiliarpsychiatrie*, 2(2), 92-97.
- Baumeister, H., Kury, S. & Bengel, J. (2011): Somatopsychische Komorbidität - Definition und Verständnis. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 54(1), 9-14.
- Beck, A.T. & Bhar, S.S. (2009): "Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis": Comment. *JAMA: Journal of the American Medical Association*, 301(9), 931.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. & Erbaugh, J. (1961): An Inventory for measuring Depression. *Archives of General Psychiatry*, 4, 561-571.
- Beckham, E.E. (1989): Improvement after evaluation in psychotherapy of depression: Evidence of a placebo effect? *Journal of Clinical Psychology*, 45(6), 945-950.
- Beierlein, V., Bergelt, C., Bullinger, M., Brähler, E. & Morfeld, M. (2010, Sept. 15 - 18). *Messung der gesundheitsbezogenen Lebensqualität mit dem SF-8: Unterschiede in den Normdaten einer schriftlich versus einer telefonisch befragten repräsentativen deutschen Bevölkerungsstichprobe*. Paper presented at the Gemeinsamer Kongress der Deutschen Gesellschaft für Medizinische Psychologie (DGMP) und der Deutschen Gesellschaft für Medizinische Soziologie (DGMS), Giessen, Germany.
- Beutler, L.E., Castonguay, L.G. & Follette, W.C. (2006): Therapeutic factors in dysphoric disorders. *Journal of Clinical Psychology*, 62(6), 639-647.
- Beutler, L.E., Malik, M., Alimohamed, S., Harwood, T.M., Talebi, H., Noble, S. & Wong, E. (2004). Therapist Variables. In Bergin and Garfield's *Handbook of Psychotherapy and Behavior Change* (5 ed.). New York: Wiley.

- Blais, M.A., Jacobo, M.C. & Smith, S.R. (2010): Exploring therapeutic alliance in brief inpatient psychotherapy: A preliminary study. *Clinical psychology & psychotherapy*, 17(5), 368-394.
- Borgart, E.J. & Meermann, R. (1999): Bedingungsfaktoren unterschiedlicher Behandlungsdauer bei Angststörungen im Rahmen stationärer Verhaltenstherapie. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 49, 109-113.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler (Vol. 6)*. Berlin: Springer.
- Broda, M., Bürger, W., Dinger-Broda, A. & Massing, H. (1996). *Die Berus-Studie. Zur Ergebnisevaluation der Therapie psychosomatischer Störungen bei gewerblichen Arbeitnehmern*. Bad Münstereifel: Westkreuz.
- Brosius, F. (2008). *SPSS 16 : das mitp-Standardwerk*. Heidelberg: mitp.
- Bryk, A.S. & Raudenbusch, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.
- Buchkremer, G. & Klingberg, S. (2001): Was ist wissenschaftlich fundierte Psychotherapie? Zur Diskussion um Leitlinien für die Psychotherapieforschung. *Nervenarzt*, 72(1), 20-30.
- Bullinger, M. (2001): Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36-Health Survey. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 43, 190-197.
- Bullinger, M. & Kirchberger, I. (1998). *SF-36. Fragebogen zum Gesundheitszustand*. Göttingen: Hogrefe.
- Cahill, J., Barkham, M., Stiles, W.B., Twigg, E., Rees, A., Hardy, G.E. & Evans, C. (2006): Convergent validity of the CORE measures with measures of depression for clients in brief cognitive therapy for depression. *Journal of Counseling Psychology*, 53(253-259).
- Castonguay, L.G. & Beutler, L.E. (2006): Principles of therapeutic change: a task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, 62(6), 631-638.
- Cicchetti, D.V. (1994): Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- Cinar, S., Oude Voshaar, R.C., Janzing, J.G.E., Birkenhager, T.K., Buitelaar, J.K. & van den Broek, W.W. (2010): The course of depressive symptoms in unipolar depressive disorder during electroconvulsive therapy: a latent class analysis. *Journal of Affective Disorders*, 124(1-2), 141-147.
- Clarkin, J.F. & Levy, K.N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 194-226). New York: John Wiley.
- Cohen, J. (1960): A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences (2 ed.)*. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cuijpers, P., van Lier, P.A., van Straten, A. & Donker, M. (2005): Examining differential effects of psychological treatment of depressive disorder: An application of trajectory analyses. *Journal of Affective Disorders*, 89(1-3), 137-146.
- Cuijpers, P., van Straten, A., Warmerdam, L. & Andersson, G. (2009): Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Depression and Anxiety*, 26(3), 279-288.
- Damke, B. & Koechel, R. (2002). Therapieabbruch in der stationären Psychosomatik I: Eine Pilotstudie. In M. Bassler (Ed.), *Stationäre Gruppenpsychotherapie. Mainzer Werkstatt ueber empirische Forschung von stationärer Psychotherapie 2001* (pp. 9-18). Giessen: Psychosozial-Verlag.
- de Jong-Meyer, R., Hautzinger, M., Rudolf, G.A.E., Strauss, W. & Frick, U. (1996): Die Überprüfung der Wirksamkeit einer Kombination von Antidepressiva- und Verhaltenstherapie bei endogen depressiven Patienten: Varianzanalytische Ergebnisse zu den Haupt- und Nebenkriterien des Therapieerfolgs. *Zeitschrift fuer Klinische Psychologie*, 25(2), 93-109.
- de Jong-Meyer, R., Hautzinger, M., Rudolf, G.A.E., Strauß, W. & Frieck, U. (1996): Prädiktions- und Verlaufsanalysen bei kombinierter psychologischer und medikamentöser Therapie endogen depressiver Patienten. *Zeitschrift für Klinische Psychologie*, 25(2), 110-129.
- Deutsche Rentenversicherung Bund. (2005). *VDR Statistik Rehabilitation des jahres 2004 (Vol. 154)*. Frankfurt am Main: DRV.
- Deutsche Rentenversicherung Bund. (2006). *Leitlinien für die sozialmedizinische Beurteilung von Menschen mit psychischen Störungen*. Berlin: Eigenverlag.
- Deutsche Rentenversicherung Bund. (2007). *Klassifikation therapeutischer Leistungen in der medizinischen Rehabilitation (KTL)*. Berlin: Deutsche Rentenversicherung Bund.
- Deutsche Rentenversicherung Bund. (2010). *Anforderungsprofil für eine stationäre Einrichtung zur medizinischen Rehabilitation von Erwachsenen mit psychosomatischen und psychischen Störungen*. Berlin: Eigenverlag.
- Deutsche Rentenversicherung Bund. (2011). *Reha-Therapiestandards Depressive Störungen - Leitlinie für die medizinische Rehabilitation der Rentenversicherung*. Berlin: Deutsche Rentenversicherung Bund.
- DGPPN, BÄK, KBV, AWMF, AkdÄ, BpTK, BApK, DAGSHG, DEGAM, DGPM, DGPs & DGRW (Hrsg) für die Leitliniengruppe Unipolare Depression. (2009). *S3-Leitlinie/Nationale Versorgungs-Leitlinie Unipolare Depression-Langfassung*. from www.dgppn.de, www.versorgungsleitlinien.de, www.awmf-leitlinien.de.
- Dilcher, K., Mestel, R., Klingelhöfer, J., Köbel, W., Sprenger, B. & Stauss, K. (2000). Psychosomatische Kliniken. In N. Gerdes, H. Weidemann & W. H. Jäckel (Eds.), *Die Protos-Studie* (pp. 173-202). Darmstadt: Steinkopff.

- Dilling, H., Mombour, W. & Schmidt, M.H. (2000). *Internationale Klassifikation psychischer Störungen, ICD-10 Kapitel V (F) Klinisch-diagnostische Leitlinien (Vol. 4)*. Bern: Hans Huber.
- Dirmaier, J., Krattenmacher, T., Watzke, B., Koch, U., Schulz, H. & Barghaan, D. (2010): Evidence-based treatments in the rehabilitation of patients with depression a literature review. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 60, 83-97.
- Dohrenbusch, R. & Scholz, O.B. (2002): Zur faktoriellen und kriterienbezogenen Validität des Fragebogens zur Psychotherapiemotivation (FPTM40). *Diagnostica*, 48(3), 150-158.
- Downs, S.H. & Black, N. (1998): The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52, 337-384.
- Drymalski, W.M. & Washburn, J.J. (2011): Sudden gains in the treatment of depression in a partial hospitalization program. *Journal of Consulting and Clinical Psychology*, 79(3), 364-368.
- Edwards, J.E., Moore, R.A. & McQuay, H.J. (2004): Individual patient meta-analysis of single-dose rofecoxib in postoperative pain. *BMC Anesthesiology*, 4(3).
- Efron, B. (1979): Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.
- Eid, M., Geiser, C. & Nußbeck, F. (2008): Neuere psychometrische Ansätze der Veränderungsmessung. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 56(3), 181-189.
- Ellert, U., Lampert, T. & Ravens-Sieberer, U. (2005): Measuring health-related quality of life with the SF-8. Normal sample of the German population. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 48(12), 1330-1337.
- Elm, E.v., Altman, D.G., Egger, M., Pocock, S.J., Gotsche, P.C., Vandenbroucke, J.P. & Initiative, S. (2007): The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, 370(9596), 1453-1457.
- Emmelkamp, P.M.G. (2004). Behavior Therapy with Adults. In M. J. Lambert (Ed.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (5 ed., pp. 393-446). New York: John Wiley.
- Emmelkamp, P.M.G., Benner, A., Kuipers, A., Feiertag, G.A., Koster, H.C. & Van Apeldoorn, F.J. (2004): Comparison of brief dynamic and cognitive-behavioural therapies in avoidant personality disorder. *British Journal of Psychiatry*, 189, 60-64.
- Erdfelder, E., Faul, F. & Buchner, A. (1996): GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Erhart, M., Wetzel, R., Krugel, A. & Ravens-Sieberer, U. (2005): Assessment of health-related quality of life with the German SF-8. A comparison of telephone and postal survey modes. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 48(12), 1322-1329.

- Eysenck, H.J. (1952): The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16(319-324).
- Falconnier, L. (2009): Socioeconomic Status in the Treatment of Depression. *American Journal of Orthopsychiatry*, 79(2), 148-158.
- Farin, E., Dudeck, A., Meffert, C., Glattacker, M., Jäckel, W.H., Beckman, U. & Polak, U. (2009): Direkte und indirekte Veränderungsmessung in der ambulanten Rehabilitation muskuloskeletaler und kardiologischer Erkrankungen. *Physikalische Medizin Rehabilitationsmedizin Kurortmedizin*, 19, 149-161.
- Franke, G. (2002). *SLC-90-R. Die Symptom Checkliste von L.R. Derogatis – Manual zur Deutschen Version*. Weinheim: Beltz.
- Guyatt, G.H., Norman, G.R., Juniper, E.F. & Griffith, L.E. (2002): A critical look at transition ratings. *Journal of Clinical Epidemiology*, 55(9), 900-908.
- Haas, E., Hill, R.D., Lambert, M.J. & Morrell, B. (2002): Do early responders maintain treatment gains? *Journal of Clinical Psychology*, 58(1157-1172).
- Hahlweg, K. (2005). Soziale Rahmenbedingungen. Die Bedeutung von Partnerschaft und Familie für die Therapie psychischer Störungen. In J. Kosfelder, J. Michalak, S. Vocks & U. Willutzki (Eds.), *Fortschritte der Psychotherapieforschung*. (pp. 115-147). Göttingen: Hogrefe.
- Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B. & von Witzleben, I. (2001): Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology*, 69(3), 375-382.
- Hamilton, M. (1960): A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry*, 23, 56-62.
- Harfst, T., Koch, U., Kurtz von Aschoff, C., Nutzinger, D.O., Rüddel, H. & Schulz, H. (2002): Entwicklung und Validierung einer Kurzform der Symptom Checklist-90-R. *DRV-Schriften*, 33, 71-73.
- Härter, M., Klesse, C., Berger, M., Bermejo, I., Bschor, T., Gensichen, J., Harfst, T., Hautzinger, M., Kuhner, C., Mundt, C., Niebling, W., Richter, R., Schulz, H. & Schneider, F. (2010): Evidence-based treatment of depression: what does the new S3- and national healthcare guideline Unipolar Depression really recommend? *Zeitschrift fuer Psychosomatische Medizin und Psychotherapie*, 56(4), 334-342.
- Härter, M., Sitta, P., Keller, F., Metzger, R., Wiegand, W., Schell, G., Stieglitz, R.-D., Wolfersdorf, M., Felsenstein, M. & Berger, M. (2004): Stationäre psychiatrisch-psychotherapeutische Depressionsbehandlung. Prozess- und Ergebnisqualität anhand eines Modellprojekts in Baden-Württemberg. *Nervenarzt*, 75(11), 1083-1091.
- Hartmann, A., Herzog, T. & Drinkmann, A. (1992): Psychotherapy of bulimia nervosa: what is effective? A meta-analysis. *Journal of Psychosomatic Research*, 36(2), 159-167.
- Hartmann, S. & Zepf, S. (2004): Einflüsse auf die Symptombesserung in der Psychotherapie bei Patienten mit unterschiedlichen Beschwerdebildern. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 54, 445-456.
- Hautzinger, M. & Bailer, M. (1993). *Allgemeine Depressionsskala (ADS). Die deutsche Version des CES-D*. Weinheim: Beltz.

- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). *Beck-Depressions-Inventar (BDI)*. Bern: Huber.
- Hautzinger, M., De Jong-Meyer, R., Treiber, R., Rudolf, G.A.E. & Thien, U. (1996a): Wirksamkeit Kognitiver Verhaltenstherapie, Pharmakotherapie und deren Kombination bei nicht-endogenen, unipolaren Depressionen. *Zeitschrift für Klinische Psychologie*, 25(2), 130-145.
- Hautzinger, M., de Jong-Meyer, R., Treiber, R., Rudolf, G.A.E., Thien, U. & Bailer, M. (1996b): Verlaufsanalysen und Prädiktoren des Therapieerfolgs bei psychologischer und pharmakologischer Therapie nicht-endogener Depressionen. *Zeitschrift für Klinische Psychologie*, 25(2), 146-154.
- Heuft, G., Hildenbrand, G. & Cuntz, U. (2010): The hospital financing frame regulation, psychiatric OPS figures and new reimbursement ordinance for psychosomatic medicine and psychotherapy as well as psychiatry and psychotherapy. *Zeitschrift fuer Psychosomatische Medizin und Psychotherapie*, 56(1), 86-105.
- Hill, C.E. & Lambert, M.J. (2004). Methodological Issues in Studying Psychotherapy Process and Outcome. In Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (5 ed., pp. 84-135). New York: John Wiley.
- Hillert, A., Maasche, B., Kretschmer, A., Ehrig, C., Schmitz, E. & Fichter, M. (1999): Psychosomatische Erkrankungen bei LehrerInnen. Sozialer Kontext, Inhalte und Perspektiven stationärer Behandlungen im Hinblick auf die Wiederherstellung der Arbeitsfähigkeit. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 49(9-10), 375-380.
- Hinz, A., Schumacher, J., Albani, C., Schmid, G. & Brähler, E. (2006): Bevölkerungsrepräsentative Normierung der Skala zur Allgemeinen Selbstwirksamkeitserwartung. *Diagnostica*, 52(1), 26-32.
- Hollon, S.D. & Beck, A.T. (2004). Cognitive and Cognitive behavioral Therapies. In M. J. Lambert (Ed.), Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (5 ed., pp. 447-492). New York: John Wiley.
- Hölzel, L., Wolff, A.v., Kriston, L. & Härter, M. (2010): Unter welchen Bedingungen bleibt eine stationäre Depressionsbehandlung erfolglos? Eine Analyse an routinemaessig erhobenen Daten von 511 Patienten. *Psychiatrische Praxis*, 37(1), 27-33.
- Howard, K.I., Kopta, S., Krause, M.S. & Orlinsky, D.E. (1986): The dose-effect relationship in psychotherapy. *American Psychologist*, 41, 159-164.
- Howard, K.I., Lueger, R.J., Maling, M.S. & Martinovic, Z. (1993): A phase model of psychotherapy outcome: causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61, 678-685.
- Hunsley, J. & Lee, C.M. (2007): Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice*, 38(1), pp.
- Ilardi, S.S. & Craighead, W.E. (1994): The role of non-specific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice*, 1, 138-156.

- Jacobi, C., Dahme, B. & Rustenbach, S. (1997): Vergleich kontrollierter Psycho- und Pharmakotherapiestudien bei Bulimia und Anorexia nervosa. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 47(9-10), 346-364.
- Jacobi, F., Klose, M. & Wittchen, H.U. (2004): Psychische Störungen in der deutschen Allgemeinbevölkerung. Inanspruchnahme von Gesundheitsleistungen und Ausfalltage. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 47(8), 736-744.
- Jacobson, N.S. & Truax, P. (1991): Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Jung, T. & Wickrama, K.A.S. (2008): An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, 2(1), 302-317.
- Kächele, H., Kordy, H. & Richard, M. (2001): Therapy amount and outcome of inpatient psychodynamic treatment of eating disorders in Germany: Data from a multicenter study. *Psychotherapy Research*, 11(3), pp.
- Kadera, S.W., Lambert, M.J. & Andrews, A.A. (1996): How much therapy is really enough? A session-by-session analysis of the psychotherapy dose-effect relationship. *Journal of Psychotherapy Practice and Research*, 5(2), 132-151.
- Kaplan, M.J., Lazoff, M., Kelly, K., Lukin, R. & Garver, D.L. (1990): Enlargement of cerebral third ventricle in psychotic patients with delayed response to neuroleptics. *Biological Psychiatry*, 27(2), 205-214.
- Katz, M.M., Tekell, J.L., Bowden, C.L., Brannan, S., Houston, J.P., Berman, N. & Frazer, A. (2004): Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. *Neuropsychopharmacology*, 29, 566-579.
- Kazis, L.E., Anderson, J.J. & Meenan, R.F. (1989): Effect sizes for interpreting changes in health status. *Medical Care*, 27(3), 178-189.
- Keller, F. & Hautzinger, M. (2007): Klassifikation von Verlaufskurven in der Depressionsbehandlung. Ein methodischer Beitrag. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(2), 83-92.
- Kelly, M.A.R., Roberts, J.E. & Ciesla, J. (2005): Sudden gains in cognitive behavioural treatment for depression: When do they occur and do they matter? *Behaviour Research and Therapy*, 43, 703-714.
- Kendall, P.C., Holmbeck, G. & Verduin, T. (2004). Methodology, Design, and Evaluation in Psychotherapy Research. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 16-43). New York: John Wiley.
- Klaghofer, R. & Brähler, E. (2001): Konstruktion und teststatische Prüfung einer Kurzform der SCL-90-R. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 49(2), 115-124.
- Koch, U. & Potreck-Rose, F. (1994). Stationäre psychosomatische Rehabilitation - ein Versorgungssystem in der Diskussion. In B. Strauß & A. E. Meyer (Eds.), *Psychoanalytische Psychosomatik - Theorie, Forschung und Praxis* (pp. 193-212). Stuttgart: Schattauer.

- Koch, U. & Schulz, H. (2003): Voraussetzungen und Strategien zur Entwicklung von Leitlinien in der Rehabilitation von Patienten mit psychischen/psychosomatischen Störungen. *Rehabilitation*, 42(2), 118-123.
- Koch, U. & Schulz, H. (2008). Weiterentwicklung in der Psychosomatischen Rehabilitation. In G. Schmid-Ott, S. Wiegand-Grefe, C. Jacobi, G. H. Paar, R. Meermann & F. Lamprecht (Eds.), *Rehabilitation in der Psychosomatik* (pp. 445-458). Stuttgart: Schattauer.
- Kohlmann, T. & Raspe, H. (1998): Zur Messung patientennaher Erfolgskriterien in der medizinischen Rehabilitation: Wie gut stimmen "indirekte" und "direkte" Methoden der Veränderungsmessung überein? *Rehabilitation*, 37(1), 30-37.
- Kösters, M., Nachtigall, C., Burlingame, G.M. & Strauß, B. (2006): A meta-analytic review of the effectiveness of inpatient group psychotherapy. *Group Dynamics: Theory, Research, and Practice*, 10(2), 146-163.
- Kösters, M. & Strauß, B. (2007): Wirksamkeit stationärer Gruppenpsychotherapie - Eine kritische Betrachtung deutschsprachiger Studien. *Gruppenpsychotherapie und Gruppendynamik*, 43(3), 181-200.
- Kriston, L., Hölzel, L. & Härter, M. (2009): "Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis": Comment. *JAMA: Journal of the American Medical Association*, 301(9), 930-931.
- Kriston, L., Melchior, H., Hergert, A., Bergelt, C., Watzke, B., Schulz, H. & Wolff, A.v. (2011): Class Evolution Tree: a graphical tool to support decisions on the number of classes in exploratory categorical latent variable modeling for rehabilitation research. *International Journal of Rehabilitation Research*, 34(2), 181-185.
- Lambert, M.J. (2005): Early Response in Psychotherapy: Further Evidence for the Importance of Common Factors Rather Than "Placebo Effects". *Journal of Clinical Psychology*, 61(7), 855-869.
- Lambert, M.J., Hansen, N.B. & Finch, A.E. (2001): Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69(159-172).
- Lambert, M.J. & Ogles, B.M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 139-193). New York: John Wiley.
- Lamprecht, F. & Schmidt, J. (1990). Das Zaubenberg-Projekt: zwischen Verzauberung und Ernüchterung. In S. Ahrens (Ed.), *Entwicklung und Perspektiven der Psychosomatik in der Bundesrepublik Deutschland* (pp. 97-115). Berlin: Springer.
- Landis, J.R. & Koch, G.G. (1977): The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lang, K. (2003). *Behandlungsabbrüche und Therapiemotivation in der stationären Rehabilitation von Patienten mit psychischen Erkrankungen - Entwicklung und empirische Überprüfung eines Vorhersagemodells*. Universität Hamburg, Hamburg.
- Lang, K., Schulz, H., Lotz-Rambaldi, W. & Koch, U. (1999): Behandlungsabbruch als nicht gelungene Inanspruchnahme. Entwicklung eines Vorhersagemodells fuer den

- Bereich der stationären psychosomatischen Rehabilitation. *Die Rehabilitation*, 38, Suppl. 2(160-166).
- Lauber, C., Lay, B. & Rössler, W. (2006): Length of first admission and treatment outcome in patients with unipolar depression. *Journal of Affective Disorders*, 93(1-3), 43-51.
- Leichsenring, F. & Rabung, S. (2008): Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA*, 300(13), 1551-1565.
- Lieberman, P.B., Wiitala, S.A., Elliott, B., McCormick, S. & Goyette, S.B. (1998): Decreasing length of stay: are there effects on outcomes of psychiatric hospitalization? *American Journal of Psychiatry*, 155(7), 905-909.
- Lieberz, K. & Ciemer, S. (2000): Background of non-completion of inpatient psychotherapy. *Psychotherapeut*, 45(5), 286-291.
- Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.
- Lo, Y., Mendell, N.R. & Rubin, D.B. (2001): Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.
- Luborsky, L., Barber, J.P., Siqueland, L., Johnson, S., Najavits, L.M., Frank, A. & Daley, D. (1996): The revised Helping Alliance questionnaire (HAQ-II): Psychometric properties. *Journal of Psychotherapy Practice and Research*, 5(3), 260-271.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007): Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103-117.
- Lueger, R.J. (1995): Ein Phasenmodell der Veränderung in der Psychotherapie. *Psychotherapeut*, 40(267-278).
- Lueger, R.J. (1998): Using feedback on patient progress to predict the outcome of psychotherapy. *Journal of Clinical Psychology*, 54(3), 383-393.
- Lutz, W. (2003): Efficacy, effectiveness, and expected treatment response in psychotherapy. *Journal of Clinical Psychology*, 59(7), 745-750.
- Lutz, W., Bachmann, F., Tschitsaz, A., Smart, D.W. & Lambert, M.J. (2007a): Zeitliche und sequenzielle Muster von nichtlinearen Veränderungen im Therapieverlauf. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(4), 261-269.
- Lutz, W., Grawe, K. & Tholen, S. (2003): Prognose des individuellen Behandlungsverlaufs für unterschiedliche klinisch-therapeutische Behandlungsmodalitäten in der Verhaltenstherapie. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 51(1), 51-70.
- Lutz, W., Rafaeli, E., Howard, K.I. & Martinovich, Z. (2002): Adaptive modelling of progress in outpatient psychotherapy. *Psychotherapy Research*, 12(4), 427-443.
- Lutz, W., Stulz, N. & Köck, K. (2009): Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, 118(1-3), 60-68.

- Lutz, W., Stulz, N., Smart, D.W. & Lambert, M.J. (2007b): Die Identifikation früher Veränderungsmuster in der ambulanten Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 36(2), 93-104.
- Maier-Riehle, B. & Zwingmann, C. (2000): Effektstärkevarianten beim Eingruppen-Pra-Post-Design: Eine kritische Betrachtung. *Die Rehabilitation*, 39(4), 189-199.
- Martin, D.J., Garske, J.P. & Davis, M.K. (2000): Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68(3), 438-450.
- Martinsohn-Schittkowski, W. & Tolzin, C.J. (2008): Verweildauer kann und sollte weiter sinken. *Psychiatrische Praxis*, 35, 267-268.
- McNeilly, C.L. & Howard, K.I. (1991): The effects of psychotherapy: A reevaluation based on dosage. *Psychotherapy Research*, 1, 74-78.
- Melchior, H., Hergert, A., Hofreuter-Gätgens, K., Bergelt, C., Morfeld, M., Schulz, H., Koch, U. & Watzke, B. (2010): Prädiktoren der stationären Behandlungsdauer von Patienten mit psychischen Störungen - Eine systematische Literaturanalyse. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 56(4), 399-418.
- Menard, S. (2010). Logistic Regression Diagnostics and Problems of Inference. In S. Menard (Ed.), *Logistic Regression: From Introductory to Advanced Concepts and Applications* (pp. 125ff). Thousand Oaks, California: Sage.
- Miller, I.W., Norman, W.H. & Keitner, G.I. (1999): Combined treatment for patients with double depression. *Psychotherapy and Psychosomatics*, 68(4), 180-185.
- Mussgay, L., Terporten, G. & Rueddel, H. (1996). *Merkmale von Abbrechern einer stationären Rehabilitationsbehandlung*. Paper presented at the 14. Klinisch-psychologisches Symposium der Fachgruppe Klinische Psychologie der Deutschen Gesellschaft für Psychologie, Bamberg.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage.
- Muthén, B. & Muthén, L. (2008). *Mplus User's Guide*. Los Angeles, CA: Muthén, Muthén.
- Neeb, K., Winkler, K., Schröder, A. & Mestel, R. (2001): Welchen Einfluss hat eine Therapiezeitverkürzung auf den Behandlungserfolg stationärer Psychotherapie? *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 14, 333-340.
- Newman, M.G. & Stiles, W.B. (2006): Therapeutic factors in treating anxiety disorders. *Journal of Clinical Psychology*, 62(6), 649-659.
- Nosper, M. (1999a): Der Erfolg Psychosomatischer Rehabilitation in Abhängigkeit von der Behandlungsdauer. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 49, 354-360.
- Nosper, M. (1999b). *Psychosomatische Rehabilitation. Ergebnis- und Prozessqualität von Einzel- und Psychotherapien*. Berlin: Logos.
- Nosper, M. (2008): Die Dauer psychosomatischer Rehabilitation – Regelungen, Einflussfaktoren und Empfehlungen. *Rehabilitation*, 47, 8-13.

- Nübling, R., Schmidt, J. & Schulz, H. (2002). FPTM. Fragebogen zur Psychotherapiemotivation. In E. Brähler, J. Schumacher & B. Strauß (Eds.), *Diagnostische Verfahren in der Psychotherapie* (pp. 141-145). Göttingen: Hogrefe.
- Nübling, R., Schulz, H., Schmidt, J., Koch, U. & Wittmann, W.W. (2006). Fragebogen zur Psychotherapiemotivation (FPTM) - Testkonstruktion und Gütekriterien. In R. Nübling, F. A. Muthny & J. Bengel (Eds.), *Reha-Motivation und Behandlungserwartung* (Vol. 47, pp. 252-270). Bern: Huber.
- Nübling, R., Steffanowski, A., Wittmann, W.W. & Schmidt, J. (2004): Strategies of measuring the outcomes of psychosomatic rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 17, 35-44.
- Orlinsky, D.E., Ronnestad, M.H. & Willutzki, U. (2004). Fifty years of psychotherapy and process-outcome research: continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 307-390). New York: John Wiley.
- Paar, G.H. & Grohmann, S. (2000): Überlegungen zu einem „Allgemeinen Modell der psychosomatischen Rehabilitation“ mit Ableitungen zur angemessenen Behandlungsintensität und erforderlichen Verweildauer. *Rehabilitation*, 39, 8-16.
- Pfaff, H., Glaeske, G., Neugebauer, E.A.M. & Schrappe, M. (2009): Memorandum III: "Methods for Health Services Research" (Part 1). *Gesundheitswesen*, 71(8-9), 505-510.
- Potreck-Rose, F. & Koch, U. (1994). *Chronifizierungsprozesse bei psychosomatischen Patienten*. Stuttgart: Schattauer.
- Prinz, U., Nutzinger, D.O., Schulz, H., Petermann, F., Braukhaus, C. & Andreas, S. (2008): Die Symptom-Checkliste-90-R und ihre Kurzversionen: Psychometrische Analysen bei Patienten mit psychischen Erkrankungen. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, 18(337-343).
- Quitkin, F.M., McGrath, P.J., Stewart, J.W., Taylor, B.P. & Klein, D.F. (1996): Can the effects of antidepressants be observed in the first two weeks of treatment? *Neuropsychopharmacology*, 15, 390-394.
- Rabung, S. (2007). *Qualitätssicherung durch (faire) Einrichtungsvergleiche? Zum Umgang mit dem Problem fehlender Werte im Kontext der einrichtungsvergleichenden Qualitätssicherung medizinischer Rehabilitation*. Universität Hamburg, Hamburg.
- Rabung, S. (2010): Was tun bei fehlenden Werten? *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 60(12), 485.
- Rabung, S., Harfst, T., Kawski, S., Koch, U., Wittchen, H.-U. & Schulz, H. (2009): Psychometrische Überprüfung einer verkürzten Version der „Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis“ (HEALTH-49). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 55, 162-179.
- Rabung, S., Harfst, T., Koch, U. & Schulz, H. (2008). Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH) – Referenzdaten zur verkürzten 49-Item-Version „HEALTH-49“ (Stand:

- 15.01.2007). Retrieved March, 2011, from <http://www.hamburger-module.de/download/health-49-normen.pdf>
- Raspe, H., Pfaff, H., Härter, M., Hart, D., Koch-Gromus, U., Schwartz, F.-W., Siegrist, J. & Wittchen, H.-U. (2010). *Versorgungsforschung in Deutschland: Stand - Perspektiven - Förderung*. Deutsche Forschungsgemeinschaft. Bonn: Wiley-VHC Verlag.
- Renaud, J., Brent, D.A., Baugher, M., Birmaher, B., Kolko, D.J. & Bridge, J. (1998): Rapid response to psychosocial treatment for adolescent depression: A two-year follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37(11), 1184-1191.
- Richter, D. (2001): Die Dauer der stationären psychiatrischen Behandlung – Eine Übersicht über Methodik, Einflussfaktoren und Auswirkungen. *Fortschritte der Neurologie-Psychiatrie*, 69, 19-31.
- Roepke, S. & Renneberg, B. (2009): "Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis": Comment. *JAMA: Journal of the American Medical Association*, 301(9), 931-932.
- Saß, H., Wittchen, H.U., Houben, I. & Zaudig, M. (2003). *Diagnostisches und Statistisches Manual psychischer Störungen - Textrevision - DSM-IV-TR*. Göttingen: Hogrefe.
- Schauenburg, H., Sammet, I. & Strack, M. (2001): Verlauf der Symptombelastung und Vorhersage des Behandlungserfolges in der stationären Psychotherapie. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 47, 380-395.
- Schiepek, G. & Cremers, S. (2003). Ressourcenorientierung und Ressourcendiagnostik in der Psychotherapie. In H. Schemmel & J. Schaller (Eds.), *Ressourcen. Ein Hand- und Lesebuch zur therapeutischen Arbeit*. (pp. 147-193). Tübingen: DGVT Deutsche Gesellschaft fuer Verhaltenstherapie
- Schmidt, J., Nübling, R. & Wittmann, W.W. (2000): Ergebnisqualität stationärer psychosomatischer Rehabilitation nach einem Jahr. Die Patientenperspektive in 5 Programmevaluationsstudien mit 2098 Patienten. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 52(32-47).
- Schmitt, J. & Di Fabio, R.P. (2005): The validity of prospective and retrospective global change criterion measures. *Archives of Physical Medicine and Rehabilitation*, 86(12), 2270-2276.
- Schmitz-Buhl, S.M., Kriebel, R. & Paar, G.H. (1999): Zeitsensitive Therapie: Zusammenhänge zwischen Therapiedauer, Therapiemotivation, Beschwerdestärke und Behandlungserfolg in der stationären psychosomatischen Rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 45, 21-27.
- Schneider, F., Härter, M., Brand, S., Sitta, P., Menke, R., Hammer-Filipiak, U., Kudling, R., Heindl, A., Herold, K., Frommberger, U., Elmer, O., Hetzel, G., Witt, G., Wolfersdorf, M., Berger, M. & Gaebel, W. (2005): Adherence to guidelines for treatment of depression in in-patients. *British Journal of Psychiatry*, 187, 462-469.
- Schulz, H., Barghaan, D., Harfst, T., Dirmaier, J., Watzke, B. & Koch, U. (2006): Versorgungsforschung in der psychosozialen Medizin. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 49(2), 175-187.

- Schulz, H., Barghaan, D., Harfst, T. & Koch, U. (2008). *Psychotherapeutische Versorgung*. Berlin: Robert Koch-Institut.
- Schulz, H., Harfst, T., Andreas, S., Kawski, S., Koch, U. & Rabung, S. (2009): Zusammenhang zwischen Patienten- und Therapeuteneinschätzungen der Ergebnisqualität in der Rehabilitation von Patienten mit psychischen/psychosomatischen Erkrankungen am Beispiel des SF-8. *Rehabilitation*, 48(5), 270-276.
- Schulz, H. & Koch, U. (2002): Zur stationären psychosomatisch-psychotherapeutischen Versorgung in Norddeutschland - Expertise zu Fragen des Bedarfs und zur Versorgungsstruktur. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 52(5), 244-247.
- Schulz, H., Lang, K., Nübling, R. & Koch, U. (2003): Psychometrische Überprüfung einer Kurzform des Fragebogens zur Psychotherapiemotivation – FPTM-23. *Diagnostica*, 49(2), 83-93.
- Schulz, H., Lotz-Rambaldi, W., Koch, U., Jürgensen, R. & Rüdgel, H. (1999): 1-Jahres-Katamnese stationärer psychosomatischer Rehabilitation nach differentieller Zuweisung zu psychoanalytisch oder verhaltenstherapeutisch orientierter Behandlung. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 49, 114-130.
- Schulz, H., Nübling, R. & Rüdgel, H. (1995): Entwicklung einer Kurzform eines Fragebogens zur Psychotherapiemotivation. *Verhaltenstherapie*, 5, 89-95.
- Schütz, A. & Wiesner, C. (2000). Partnerschaft und Gesundheitszustand. In P. Kaiser (Ed.), *Partnerschaft und Paartherapie*. (pp. 193-218). Göttingen: Hogrefe.
- Schwarzer, G., Türp, J.C. & Antes, G. (2004): Das Odds Ratio in Interventionsstudien. *Deutsche Zahnärztliche Zeitschrift, EbM-Splitter*, 59(10), 549-550.
- Schwarzer, R. & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, UK: NFER-NELSON.
- Shadish, W.R., Matt, G.E., Navarro, A.M. & Phillips, G. (2000): The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological Bulletin*, 126(4), 512-529.
- Sharf, J., Primavera, L.H. & Diener, M.J. (2010): Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy Theory, Research, Practice, Training*, 47(4), 637-645.
- Spießl, H. & Klein, H.E. (2008): Eine untere Grenze der Verweildauer ist erreicht! *Psychiatrische Praxis*, 35, 269-270.
- Statistisches Bundesamt. (2006). *Fachserie 12/Reihe 6.1 - Grunddaten der Krankenhäuser und Vorsorge- oder Rehabilitationseinrichtungen 2004*. Wiesbaden.
- Steffanowski, A., Löschmann, C., Schmidt, J., Wittmann, W.W. & Nübling, R. (2007). *Meta-Analyse der Effekte stationärer psychosomatischer Rehabilitation (Vol. 48)*. Bern: Huber.

- Stevens, J.P. (2009). *Applied multivariate statistics for the social sciences (5 ed.)*. New York, NY: Routledge.
- Stewart, J.W., Quitkin, F.M., McGrath, P.J., Amsterdam, J., Fava, M., Fawcett, J., Reimherr, F., Roscnbaum, J., Beasley, C. & Roback, P. (1998): Use of pattern analysis to predict differential relapse of remitted patients with major depression during 1 year of treatment with fluoxetine or placebo. *Archives of General Psychiatry*, 55(4), 334-343.
- Stieglitz, R.-D. & Baumann, U. (2001). Veränderungsmessung. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Eds.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie (Vol. 2, pp. 21-38)*. Stuttgart: Thieme.
- Stiles, W.B., Barkham, M., Connell, J. & Mellor-Clark, J. (2008a): Responsive Regulation of Treatment Duration in Routine Practice. *Journal of Consulting and Clinical Psychology*, 76(2), 298-305.
- Stiles, W.B., Barkham, M., Iveson, S., Iveson, M., Leach, C., Lucock, M., Shapiro, D.A. & Hardy, G.E. (2003): Early sudden gains under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, 71, 14-21.
- Stiles, W.B., Barkham, M., Mellor-Clark, J. & Connell, J. (2008b): Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: replication in a larger sample. *Psychological Medicine*, 38(5), 677-688.
- Stiles, W.B., Honos-Webb, L. & Surko, M. (1998): Responsiveness in Psychotherapy. *Clinical Psychology: Science and Practice*, 5(N4), 439-458.
- Stirman, S.W., Derubeis, R.J., Crits-Christoph, P. & Rothman, A. (2005): Can the randomized controlled trial literature generalize to nonrandomized patients? *Journal of Consulting and Clinical Psychology*, 73(1), 127-135.
- Strauss, B. & Wittmann, W.W. (2005). Psychotherapieforschung: Grundlagen und Ergebnisse. In W. Senf & M. Broda (Eds.), *Praxis der Psychotherapie (Vol. 3, pp. 760-781)*. Stuttgart: Thieme.
- Streeck, U. (2005). Psychodynamische Therapieverfahren. In W. Senf & M. Broda (Eds.), *Praxis der Psychotherapie (Vol. 3, pp. 238-258)*. Stuttgart: Thieme.
- Stulz, N., Lutz, W., Leach, C., Lucock, M. & Barkham, M. (2007): Shapes of Early Change in Psychotherapy Under Routine Outpatient Conditions. *Journal of Consulting and Clinical Psychology*, 75(6), 864-874.
- Stulz, N., Thase, M.E., Klein, D.N., Manber, R. & Crits-Christoph, P. (2010): Differential Effects of Treatments for Chronic Depression: A Latent Growth Model Reanalysis. *Journal of Consulting and Clinical Psychology*, 78(3), 409-419.
- Tadic, A., Helmreich, I., Mergl, R., Hautzinger, M., Kohnen, R., Henkel, V. & Hegerl, U. (2010): Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders*, 120(1-3), 86-93.
- Talala, K., Huurre, T., Aro, H., Martelin, T. & Prattala, R. (2009): Trends in socio-economic differences in self-reported depression during the years 1979-2002 in Finland. *Social Psychiatry and Psychiatric Epidemiology*, 44(10), 871-879.

- Tang, T.Z. & DeRubeis, R.J. (1999a): Reconsidering rapid early response in cognitive behavioral therapy for depression. *Clinical Psychology: Science and Practice*, 6(3), 283-288.
- Tang, T.Z. & DeRubeis, R.J. (1999b): Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67, 894-904.
- Tang, T.Z., DeRubeis, R.J., Hollon, S.D., Amsterdam, J. & Shelton, R. (2007): Sudden Gains in Cognitive Therapy of Depression and Depression Relapse/Recurrence. *Journal of Consulting and Clinical Psychology*, 75(3), 404-408.
- Tang, T.Z., Luborsky, L. & Andrusyna, T. (2002): Sudden gains in recovering from depression: are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of Consulting and Clinical Psychology*, 70(2), 444-447.
- Tedlow, J., Fava, M., Uebelacker, L., Nierenberg, A.A., Alpert, J.E. & Rosenbaum, J. (1998): Outcome definitions and predictors in depression. *Psychotherapy and Psychosomatics*, 67(4-5), 266-270.
- Teusch, L., Bohme, H., Finke, J., Gastpar, M. & Skerra, B. (2003): Antidepressant medication and the assimilation of problematic experiences in psychotherapy. *Psychotherapy Research*, 13(3), 307-322.
- Thase, M.E. & Jindal, R.D. (2004). Combining Psychotherapy and Psychopharmacology for Treatment of Mental Disorders. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 743-765). New York: John Wiley.
- Thompson, M.G., Thompson, L. & Gallagher-Thompson, D. (1995): Linear and nonlinear changes in mood between psychotherapy sessions: Implications for treatment outcome and relapse risk. *Psychotherapy Research*, 5, 327-336.
- Tritt, K., von Heymann, F., Loew, T.H., Benker, B., Bleichner, F., Buchmueller, R., Findeisen, P., Galuska, J., Kalleder, W., Lettner, F., Michelitsch, B., Pfitzer, F., Stadtmueller, G. & Zaudig, M. (2003): Patienten in stationärer psychosomatischer Krankenhausbehandlung: Patientencharakteristika und Behandlungsergebnisse anhand der Psy-Bado-PTM. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 8(1), 244-251.
- Trivedi, M.H., Morris, D.W., Grannemann, B.D. & Mahadi, S. (2005): Symptom clusters as predictors of late response to antidepressant treatment. *Journal of Clinical Psychiatry*, 66(8), 1064-1070.
- Truant, G.S. (1999): Assessment of suitability for psychotherapy: II. Assessment based on basic process goals. *American Journal of Psychotherapy*, 53(1), 17-34.
- Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., Petrovic, A., Zobel, A., Henigsberg, N., Rietschel, M., Aitchison, K.J., Farmer, A. & McGuffin, P. (2010): Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine*, 40(8), 1367-1377.
- van Calker, D., Zobel, I., Dykieriek, P., Deimel, C.M., Kech, S., Lieb, K., Berger, M. & Schramm, E. (2009): Time course of response to antidepressants: Predictive value of early improvement and effect of additional psychotherapy. *Journal of Affective Disorders*, 114, 243-263.

- Van, H.L., Dekker, J., Koelen, J., Kool, S., van Aalst, G., Hendriksen, M., Peen, J. & Schoevers, R. (2009): Patient preference compared with random allocation in short-term psychodynamic supportive psychotherapy with indicated addition of pharmacotherapy for depression. *Psychotherapy Research*, 19(2), 205-212.
- Van, H.L., Schoevers, R.A., Kool, S., Hendriksen, M., Peen, J. & Dekker, J. (2008): Does early response predict outcome in psychotherapy and combined therapy for major depression? *Journal of Affective Disorders*, 105(1-3), 261-265.
- van Stel, H.F., Maille, A.R., Colland, V.T. & Everaerd, W. (2003): Interpretation of change and longitudinal validity of the quality of life for respiratory illness questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. *Quality of Life Research*, 12(2), 133-145.
- Verband Deutscher Rentenversicherungsträger. (2004). *VDR Statistik Rehabilitation des Jahres 2004 (Vol. 154)*. Berlin: Eigenverlag.
- von Heymann, F., Zaudig, M. & Tritt, K. (2003): Die diagnosebezogene Behandlungsdauer in der Psychosomatischen und Psychotherapeutischen Medizin: Eine homogene Größe? – Erste Ergebnisse der Multicenter-Basisdokumentation (Psy-BaDo-PTM) als Grundlage qualitätssichernder Maßnahmen in der stationären Psychosomatik. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 62, 209-221.
- Wagnild, G.M. & Young, H.M. (1993): Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement*, 1(2), 165-178.
- Ware, J.E. (1993). *SF-36 Health Survey: Manual and Interpretation Guide*: The Health Institute, New England Medical Center.
- Ware, J.E. (2000): SF-36 Health survey update. *Spine*, 25, 3130-3139.
- Ware, J.E., Kosinski, M., Dewey, J.E. & Gandek, B. (2001). *How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8™ Health Survey*. Lincoln (RI): QualityMetric Incorporated.
- Watzke, B., Barghaan, D., Harfst, T., Koch, U. & Schulz, H. (2009). Versorgungsforschung. In J. Margraf & S. Schneider (Eds.), *Lehrbuch der Verhaltenstherapie*. Band 1: Grundlagen, Diagnostik, Verfahren, Rahmenbedingungen (pp. 845-856). Heidelberg: Springer.
- Watzke, B., Büscher, C., Koch, U. & Schulz, H. (2005): Rehabilitative Elemente bei der Behandlung von Menschen mit psychischen Störungen: Werden sie bei der Entwicklung von Leitlinien ausreichend berücksichtigt? *Psychiatrische Praxis*, 32(2), 79-86.
- Watzke, B., Rüdell, H., Jürgensen, R., Koch, U., Kriston, L., Grothgar, B. & Schulz, H. (2010): Effectiveness of systematic treatment selection for psychodynamic and cognitive-behavioural therapy: Randomised controlled trial in routine mental healthcare. *British Journal of Psychiatry*, 197(2), 96-105.
- Wilson, G.T. (1999): Commentary. Rapid response to cognitive behavior therapy. *Clinical Psychology: Science and Practice*, 6(3), 289-292.

- Wirtz, M. (2004): Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Rehabilitation*, 43, 109-115.
- Wirtz, M. & Kutschmann, M. (2007): Analyzing interrater agreement for categorical data using Cohen's kappa and alternative coefficients. *Rehabilitation*, 46(6), 370-377.
- Wittchen, H.-U. & Jacobi, F. (2001): Die Versorgungssituation psychischer Störungen in Deutschland. Eine klinisch epidemiologische Abschätzung anhand des Bundes-Gesundheitssurveys 1998. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 44, 993-1000.
- Zeeck, A., Scheidt, C.E., Hartmann, A. & Wirsching, M. (2003): Stationäre oder teilstationäre Psychotherapie? *Psychotherapeut*, 48(6), 420-425.
- Zielke, M. (1993). *Wirksamkeit stationärer Verhaltenstherapie*. Weinheim: Psychologie Verlags Union.
- Zielke, M. (2006): Entwicklung differentieller Behandlungszeitkorridore für die stationäre medizinische Rehabilitation auf der Basis einer Schweregradstratifizierung von psychischen und psychosomatischen Erkrankungen. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 74, 301-308.
- Zielke, M. (2008). *Kosten-Nutzen-Relation der Psychosomatischen Rehabilitation aus gesundheitsökonomischer Perspektive*. Stuttgart: Schattauer.
- Zielke, M., Dehmlow, A., Wülbeck, B. & Limbacher, K. (1997): Einflussfaktoren auf die Behandlungsdauer bei psychosomatischen Erkrankungen in der stationären Verhaltenstherapie. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 10, 22-56.
- Zwingmann, C. & Wirtz, M. (2005): Regression to the mean. *Rehabilitation*, 44(4), 244-251.

Abbildungsverzeichnis

Abbildung 1:	Studiendesign	59
Abbildung 2:	Exemplarische Verlaufskurven für unterschiedliche Ausprägungen der Kurvenparameter Intercept, linearer und quadratischer Slope	78
Abbildung 3:	Statistische Poweranalyse für multiple hierarchische Regression.....	87
Abbildung 4:	Patientenflussdiagramm, Repräsentativitäts- und Drop-Out-Analysen	90
Abbildung 5:	Verteilung der Erstdiagnosen der Teilnehmer und der Nicht-Teilnehmer	92
Abbildung 6:	Verteilung der ersten Entlassdiagnosen über die Analyse- und Drop-Out-Stichproben.....	106
Abbildung 7:	Graphische Darstellung des BIC für 1- bis 6-Klassenlösungen.....	108
Abbildung 8:	Graphische Darstellung der Log Likelihood-Werte für 1- bis 6-Klassenlösungen	109
Abbildung 9:	Class Evolution Tree – Verteilung und Patientenfluss über die sechs modellierten Klassenlösungen (modifiziert nach Kriston et al., 2011)	111
Abbildung 10:	Verlaufskurven der sechs Verlaufsklassen.....	112
Abbildung 11:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 1 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 9$)	114
Abbildung 12:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 2 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 3$)	114
Abbildung 13:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 3 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 69$)	115
Abbildung 14:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 4 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 418$)	115
Abbildung 15:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 5 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 55$)	116
Abbildung 16:	6-Klassen-Lösung: modellierte Verlaufskurve der Klasse 6 und beobachtete Verläufe aller zu der Klasse zugeordneten Patienten ($n = 22$)	116
Abbildung 17:	Häufigkeitsverteilung der explorativen Verlaufgruppen.....	119

Abbildung 18:	Häufigkeitsverteilung der theoriegeleiteten Verlaufgruppen	120
Abbildung 19:	Graphische Darstellung der Kreuztabelle der Verteilung der theoriegeleiteten und explorativen Verlaufgruppen in den jeweiligen anderen Gruppen	124
Abbildung 20:	Wöchentliche Verbesserungsraten nach den Kriterien (1) Verbesserung mit mindestens mittlerer Effektstärke; (2) Verbesserung um mindestens 50 % der Ausgangsbelastung; (3) Verbesserung mit mindestens klinischer Signifikanz	129
Abbildung 21:	Verteilung der explorativen Verlaufgruppen in den Diagnosegruppen	132
Abbildung 22:	Verteilung der Diagnosegruppen in den explorativen Verlaufgruppen	133
Abbildung 23:	Verteilung der theoriegeleiteten Verlaufgruppen in den Diagnosegruppen	137
Abbildung 24:	Verteilung der Diagnosegruppen in den theoriegeleiteten Verlaufgruppen	138
Abbildung 25:	HEALTH-49-A-Mittelwerte der explorativen Verlaufgruppen zur Aufnahme und Katamnese	157
Abbildung 26:	HEALTH-49-A-Mittelwerte der explorativen Verlaufgruppen zur Aufnahme, Entlassung und Katamnese	157
Abbildung 27:	HEALTH-49-A-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme und Katamnese	164
Abbildung 28:	HEALTH-49-A-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme, Entlassung und Katamnese	165
Abbildung 29:	SF-8-Psych.-Mittelwerte der explorativen Verlaufgruppen zur Aufnahme und Katamnese	171
Abbildung 30:	SF-8-Psych.-Mittelwerte der explorativen Verlaufgruppen zur Aufnahme, Entlassung und Katamnese	171
Abbildung 31:	SF-8-Psych.-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme und Katamnese	177
Abbildung 32:	SF-8-Psych.-Mittelwerte der theoriegeleiteten Verlaufgruppen zur Aufnahme, Entlassung und Katamnese	178

Tabellenverzeichnis

Tabelle 1:	Schematische Darstellung der theoriegeleiteten Gruppenbildung.....	72
Tabelle 2:	Vierfeldertafel zum Erfolg und Misserfolg zweier Therapiebedingungen.....	82
Tabelle 3:	Verwendete Effektstärkemaße und ihre Interpretation	84
Tabelle 4:	Vergleich der Teilnehmer und der Nicht-Teilnehmer	91
Tabelle 5:	Soziodemographische Patientenmerkmale der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe.....	93
Tabelle 6:	Sozialmedizinische und klinische Patientenmerkmale der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe.....	96
Tabelle 7:	Diagnosenverteilung der Verlaufsstichprobe sowie Vergleiche mit der Drop-Out-Stichprobe	98
Tabelle 8:	Vergleich der Drop-Out-Raten im Verlauf zwischen den drei Kliniken.....	99
Tabelle 9:	Soziodemographische Patientenmerkmale: Vergleiche der Verlaufs- Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe	100
Tabelle 10:	Sozialmedizinische und klinische Patientenmerkmale: Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out- Stichprobe.....	102
Tabelle 11:	Diagnoseverteilung: Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe	103
Tabelle 12:	Vergleich der Drop-Out-Raten zur Katamnese zwischen den drei Kliniken	107
Tabelle 13:	Klassengrößen und Modellgütekriterien der Klassenlösungen 1 - 6 der Initialphase und 6-Klassenlösung der Validierungsphase	108
Tabelle 14:	Kurvenparameter der sechs Verlaufskassen.....	112
Tabelle 15:	Patienten mit einer maximalen Klassenzuordnungswahrscheinlichkeit von weniger als 50%.....	117
Tabelle 16:	Inhaltlich modifizierte Klassen: Verlaufsgruppen	118
Tabelle 17:	Absolute und relative Häufigkeit der explorativen Verlaufsgruppen	118
Tabelle 18:	Absolute und relative Häufigkeit der theoriegeleiteten Verlaufsgruppen	120
Tabelle 19:	Schematische Darstellung der theoriegeleiteten Gruppenbildung mit den jeweiligen Gruppengrößen.....	121

Tabelle 20:	Kreuztabelle der Verteilung der theoriegeleiteten und explorativen Verlaufsgruppen in den jeweiligen anderen Gruppen.....	123
Tabelle 21:	Zusammenhang zwischen den explorativen und den theoriegeleiteten Verlaufsgruppen.....	125
Tabelle 22:	Verteilung bis zum Erreichen des Verbesserungskriteriums: <i>Verbesserung mit einer mittleren Effektstärke (n = 576)</i>	126
Tabelle 23:	Verteilung bis zum Erreichen des Verbesserungskriteriums: <i>Verbesserung um 50 Prozent der Ausgangsbelastung (n = 576)</i>	127
Tabelle 24:	Verteilung bis zum Erreichen des Verbesserungskriteriums: <i>Klinisch signifikante Verbesserung (n = 464)</i>	128
Tabelle 25:	Zusammenhang zwischen der Ausgangsbelastung (HEALTH-49 A zu t0) und Verlaufsgruppen	141
Tabelle 26:	Auswahl der Prädiktoren zur Vorhersage der Verlaufsgruppen.....	143
Tabelle 27:	Modellzusammenfassung der logistischen Regression zur Vorhersage der Gruppenzugehörigkeit zu den <i>explorativen Verlaufsgruppen</i>	146
Tabelle 28:	Signifikante Prädiktoren der logistische Regression; Zielkriterium <i>explorative Verlaufsgruppen</i>	147
Tabelle 29:	Odds Ratios, 95% Konfidenzintervall und Signifikanz der logistischen Regression; Zielkriterium <i>explorative Verlaufsgruppen</i> (Referenzgruppe: Linear Response _{EG} , n = 367)	148
Tabelle 30:	Modellzusammenfassung der logistischen Regression zur Vorhersage der Gruppenzugehörigkeit zu den <i>theoriegeleiteten Verlaufsgruppen</i>	151
Tabelle 31:	Signifikante Prädiktoren der logistische Regression; Zielkriterium <i>theoriegeleitete Verlaufsgruppen</i>	152
Tabelle 32:	Odds Ratios, 95% Konfidenzintervall und Signifikanz der logistischen Regression; Zielkriterium <i>theoriegeleitete Verlaufsgruppen</i> (Referenzgruppe: Early Response _{TG} , n = 259).....	153
Tabelle 33:	Ergebnisse der Aufnahme-Katamnese-Vergleiche der <i>Symptombelastung (HEALTH-49 A)</i> zwischen den <i>explorativen Verlaufsgruppen</i> (ANCOVA; n = 424).....	156
Tabelle 34:	Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der <i>Symptombelastung (HEALTH-49 A)</i> zwischen jeweils zwei <i>explorativen Verlaufsgruppen</i> (ANCOVAs).....	158
Tabelle 35:	Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Symptombelastung (HEALTH-49 A zu t2)</i> durch die <i>explorativen Verlaufsgruppen</i> (n = 348)	160

Tabelle 36:	Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage des <i>katamnestischen Symptombelastung (HEALTH-49 A zu t2)</i> durch die <i>explorativen Verlaufsguppen (n = 348)</i> 161	161
Tabelle 37:	Ergebnisse der Aufnahme-Katamnese-Vergleiche der <i>Symptombelastung (HEALTH-49 A)</i> zwischen den <i>theoriegeleiteten Verlaufsguppen (ANCOVA; n = 424)</i> 163	163
Tabelle 38:	Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der <i>Symptombelastung (HEALTH-49 A)</i> zwischen jeweils zwei <i>theoriegeleiteten Verlaufsguppen (ANCOVAs)</i> 165	165
Tabelle 39:	Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Symptombelastung (HEALTH-49 A zu t2)</i> durch die <i>theoriegeleiteten Verlaufsguppen (n = 348)</i> 167	167
Tabelle 40:	Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Symptombelastung (HEALTH-49 A zu t2)</i> durch die <i>theoriegeleiteten Verlaufsguppen (n = 348)</i> 168	168
Tabelle 41:	Ergebnisse der Aufnahme-Katamnese-Vergleiche der <i>gesundheitsbezogenen Lebensqualität (SF-8 Psych.)</i> zwischen den <i>explorativen Verlaufsguppen (ANCOVA; n = 412)</i> 170	170
Tabelle 42:	Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der <i>gesundheitsbezogenen Lebensqualität (SF-8 Psych.)</i> zwischen jeweils zwei <i>explorativen Verlaufsguppen (ANCOVAs)</i> 172	172
Tabelle 43:	Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Lebensqualität (SF-8 Psych. zu t2)</i> durch die <i>explorativen Verlaufsguppen (n = 362)</i> 174	174
Tabelle 44:	Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Lebensqualität (SF-8 Psych. zu t2)</i> durch die <i>explorativen Verlaufsguppen (n = 362)</i> 175	175
Tabelle 45:	Ergebnisse der Aufnahme-Katamnese-Vergleiche der <i>gesundheitsbezogenen Lebensqualität (SF-8 Psych.)</i> zwischen den <i>theoriegeleiteten Verlaufsguppen (ANCOVA; n = 412)</i> 176	176
Tabelle 46:	Ergebnisse der paarweisen Aufnahme-Katamnese-Vergleiche der <i>gesundheitsbezogenen Lebensqualität (SF-8 Psych.)</i> zwischen jeweils zwei <i>theoriegeleiteten Verlaufsguppen (ANCOVAs)</i> 178	178
Tabelle 47:	Modellzusammenfassung der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Lebensqualität (SF-8 Psych. zu t2)</i> durch die <i>theoriegeleiteten Verlaufsguppen (n = 362)</i> 179	179
Tabelle 48:	Parameterschätzer der hierarchischen Regressionsanalyse zur Vorhersage der <i>katamnestischen Lebensqualität (SF-8 Psych. zu t2)</i> durch die <i>theoriegeleiteten Verlaufsguppen (n = 362)</i> 181	181
Tabelle 49:	Vergleich der <i>explorativen</i> und <i>theoriegeleiteten Verlaufsguppen</i> hinsichtlich des Zuwachses an aufgeklärter Varianz im <i>primären</i> und <i>sekundären Outcome</i> zu t2..... 182	182

Tabelle A.1:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 1	269
Tabelle A.2:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 2	271
Tabelle A.3:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 3	272
Tabelle A.4:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 4	273
Tabelle A.5:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 5	274
Tabelle A.6:	Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 6	274
Tabelle B.1:	Soziodemographische Patientenmerkmale – Vergleiche der Verlaufsstichprobe mit der Drop-Out-Stichprobe	275
Tabelle B.2:	Soziodemographische Patientenmerkmale – Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe.....	277
Tabelle C.1:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in soziodemographischen Patientenmerkmalen	279
Tabelle C.2:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in sozialmedizinischen Patientenmerkmalen	282
Tabelle C.3:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in Prozessmerkmalen	283
Tabelle C.4:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in klinischen Patientenmerkmalen	284
Tabelle C.5:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in personenbezogenen Merkmalen	285
Tabelle C.6:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in personenbezogenen Prozessmerkmalen	287
Tabelle C.7:	Univariate Analysen: Vergleich der explorativen Verlaufgruppen in soziodemographischen dichotomen Patientenmerkmalen	288
Tabelle D.1:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in soziodemographischen Patientenmerkmalen.....	289
Tabelle D.2:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in sozialmedizinischen Patientenmerkmalen.....	292
Tabelle D.3:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in Prozessmerkmalen	293

Tabelle D.4:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in klinischen Patientenmerkmalen	294
Tabelle D.5:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in personenbezogenen Merkmalen	295
Tabelle D.6:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in personenbezogenen Prozessmerkmalen	297
Tabelle D.7:	Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in soziodemographischen dichotomen Patientenmerkmalen	298

Anhang

Anhang A: Zusammenhangsmaße der untersuchten Merkmale

Anhang B: Stichprobenbeschreibungen und Drop-Out-Analysen

Anhang C: Vergleiche der explorativen Verlaufsguppen

Anhang D: Vergleiche der theoriegeleiteten Verlaufsguppen

Anhang E: Patienteninformation, Einverständniserklärung, Katamnese-Anschreiben

Anhang A: Zusammenhangsmaße der untersuchten Merkmale

Tabelle A.1: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 1

	Geschl.	Partner	Schulb.	Berufss.	Erwerbs.	Ambulant	Stationär
Geschlecht	1.00						
In Partnerschaft	.089*	1.00					
Schulbildung	-.134**	.071	1.00				
Berufssituation	.151***	-.022	.280***	1.00			
Erwerbsstatus	.101*	.178***	.183***	.247***	1.00		
Ambulante VB	-.069	-.051	.022	-.001	.047	1.00	
Stationäre VB	.030	-.106*	-.024	-.025	-.119**	-.008	1.00
Depression	-.071	.004	.016	.041	-.050	.103*	.009
Angststörung	-.040	.049	.009	-.046	-.020	.038	.004
Anpassungsst.	.084*	.033	-.059	-.013	.095*	-.049	-.004
PTBS	-.037	-.072	-.066	-.030	-.105*	.023	.124**
Somatoforme St.	-.076	.051	-.051	-.038	-.032	.008	.029
Essstörung	-.120**	-.071	-.062	-.077	-.141**	-.009	.050
Persönlichkeitss.	.049	-.035	-.073	-.057	-.139**	.052	.079
F Sonstige	.087*	-.039	-.058	-.048	-.039	.028	-.034
Adipositas	.008	.011	-.056	-.031	-.087*	-.121**	.087*
Komorbidität	-.056	-.035	-.095*	-.060	-.137**	.036	.090*
Klinik	.119*	-.106*	.154**	.101	.161**	.106*	.061
HEALTH-49 A t0	-.144**	-.074	-.130**	-.165***	-.209***	.104*	.114**
Alter	.080	.066	-.032	.163***	.111**	-.101*	.022
BD	.042	-.094*	-.042	.037	.031	.045	.023
Dosis PT	-.156***	-.009	.064	.028	.010	.002	.028
Anteil VT	-.126**	.046	.047	-.020	.052	-.032	.064
AU-Tage	-.005	-.067	.000	-.070	-.241***	.054	.093*
HAQ	-.082*	.034	-.019	-.005	.018	.013	-.003
SWE	.102*	.036	.063	.153***	.145**	-.160***	-.065
SF-8 Psych. t0	.096*	.113**	.055	.062	.081	-.127**	-.050
FPTM LD	-.126**	-.129**	-.042	-.082	-.110**	.212***	.063
FPTM HO	-.174**	.041	.054	.011	-.002	-.029	-.002
FPTM VH	.042	-.005	-.066	-.003	-.039	.013	-.003
FPTM SZ	.011	.108*	.027	-.074	.022	.025	.008
Explorative VG	.106	.029	.201***	.182***	.214***	.114	.132*
Theoriegel. VG	.109	.060	.043	.038	.080	.040	.056
HEALTH-49 A t2	-.021	-.074	-.178***	-.193***	-.230***	.041	.154**
SF-8 Psych. t2	.055	.014	.169***	.214***	.144**	.011	-.137**

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers V; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

Die Diagnosen sind nicht ausschließlich Erstdiagnosen, sondern beziehen sich auf die aus den ersten fünf Diagnosestellen ermittelten Diagnosegruppen.

Ambulante VB: Ambulante Vorbehandlung; Stationäre VB: Stationäre Vorbehandlung; PTBS: Posttraumatische Belastungsstörung; Komorbidität mehr als eine F-Diagnose; BD: Behandlungsdauer; Dosis Psychotherapie: Dosis psychotherapienaher Maßnahmen; Anteil VT: Anteilige Dosis an verhaltenstherapeutischen Maßnahmen; AU-Tage: Arbeitsunfähigkeitstage; FPTM LD: FPTM psychischer Leidensdruck; FPTM HO: FPTM Hoffnung; FPTM VH: FPTM Verneinung psychischer Hilfsbedürftigkeit; FPTM SZ: FPTM Symptombezogene Zuwendung durch Andere; Explorative VG: Explorative Verlaufsgruppen; Theoriegel. VG: Theoriegeleitete Verlaufsgruppen.

Tabelle A.2: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 2

	Depress.	Angst	Anpass.	PTBS	Somatof.	Essst.	Persönl.
Depression	1.00						
Angststörung	.012	1.00					
Anpassungsst.	-.656***	-.140**	1.00				
PTBS	.035	.026	-.112**	1.00			
Somatoforme St.	-.007	.092*	-.060	-.031	1.00		
Essstörung	-.057	-.080	-.150**	.078	-.075	1.00	
Persönlichkeitss.	.042	-.049	-.065	.129**	-.044	.115**	1.00
F Sonstige	.095*	.045	-.096*	.016	-.032	-.004	.167***
Adipositas	-.049	.074	-.116**	.014	-.073	.113**	.041
Komorbidität	.235***	.370***	-.139**	.183***	.283***	.195***	.232***
Klinik	.227***	.106*	.262***	.114*	.114*	.345***	.168***
HEALTH-49 A t0	.226***	.146***	-.169***	.148***	.128**	.066	.100*
Alter	.042	-.038	.147***	-.040	.055	-.313***	-.125**
BD	.083*	.033	-.058	.111**	-.010	.020	-.043
Dosis PT	-.078	.049	.021	.134***	-.004	.190***	-.120**
Anteil VT	-.090*	.038	.085	.001	-.057	-.034	-.051
AU-Tage	.110*	-.016	-.033	.028	.019	-.070	-.006
HAQ	.033	-.090*	-.027	-.003	.081	-.003	-.080
SWE	-.193***	-.143**	.161***	-.104*	.020	-.058	-.088*
SF-8 Psych. t0	-.273***	-.064	.131**	-.112**	-.056	-.043	-.068
FPTM LD	.335***	.117**	-.171***	.110**	.044	.017	.074
FPTM HO	-.064	-.072	.035	.012	.040	.033	-.005
FPTM VH	-.060	-.020	.020	-.052	-.040	.027	.035
FPTM SZ	.005	.005	-.034	-.004	-.037	.055	-.053
Explorative VG	.117	.151*	.117	.137*	.178**	.165**	.186***
Theoriegel. VG	.186***	.079	.138*	.070	.037	.096	.060
HEALTH-49 A t2	.141**	.062	-.172***	.108*	.080	.053	.081
SF-8 Psych. t2	-.015	-.029	.085	-.022	-.109*	.000	-.016

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers V; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

Die Diagnosen sind nicht ausschließlich Erstdiagnosen, sondern beziehen sich auf die aus den ersten fünf Diagnosestellen ermittelten Diagnosegruppen.

PTBS: Posttraumatische Belastungsstörung; Komorbidität mehr als eine F-Diagnose; BD: Behandlungsdauer; Dosis Psychotherapie: Dosis psychotherapienaher Maßnahmen; Anteil VT: Anteilige Dosis an verhaltenstherapeutischen Maßnahmen; AU-Tage: Arbeitsunfähigkeitstage; FPTM LD: FPTM *psychischer Leidensdruck*; FPTM HO: FPTM *Hoffnung*; FPTM VH: FPTM *Verneinung psychischer Hilfsbedürftigkeit*; FPTM SZ: FPTM *Symptombezogene Zuwendung durch Andere*; Explorative VG: Explorative Verlaufsgruppen; Theoriegel. VG: Theoriegeleitete Verlaufsgruppen.

Tabelle A.3: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 3

	F Sonstige	Adipositas	Komorbidität	Klinik
F Sonstige	1.00			
Adipositas	.001	1.00		
Komorbidität	.338***	.353***	1.00	
Klinik	.188***	.214***	.228***	1.00
HEALTH-49 A t0	.090*	.053	.251***	.183***
Alter	-.044	-.004	-.099*	.298***
BD	.057	-.078	.054	.159***
Dosis PT	-.102	-.006	.079	.590***
Anteil VT	-.044	.051	-.029	.398***
AU-Tage	.045	-.043	.035	.116
HAQ	-.048	-.071	-.071	.032
SWE	-.113**	.009	-.164***	.158**
SF-8 Psych. t0	-.081	.123**	-.172***	.141**
FPTM LD	.117**	-.121**	.196***	.168***
FPTM HO	-.139**	-.004	-.050	.061
FPTM VH	.041	.058	.037	.035
FPTM SZ	.037	-.015	-.006	.082
Explorative VG	.135*	.122	.275***	.164
Theoriegel. VG	.095	.052	.117*	.213***
HEALTH-49 A t2	.129**	.123*	.151**	.183
SF-8 Psych. t2	-.048	-.197***	-.115*	.011

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (***) und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers *V*; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

Die Diagnosen sind nicht ausschließlich Erstdiagnosen, sondern beziehen sich auf die aus den ersten fünf Diagnosestellen ermittelten Diagnosegruppen.

BD: Behandlungsdauer; Dosis Psychotherapie: Dosis psychotherapienaher Maßnahmen; Anteil VT: Anteilige Dosis an verhaltenstherapeutischen Maßnahmen; AU-Tage: Arbeitsunfähigkeitstage; FPTM LD: FPTM *psychischer Leidensdruck*; FPTM HO: FPTM *Hoffnung*; FPTM VH: FPTM *Verneinung psychischer Hilfsbedürftigkeit*; FPTM SZ: FPTM *Symptombezogene Zuwendung durch Andere*; Explorative VG: Explorative Verlaufsgruppen; Theoriegel. VG: Theoriegeleitete Verlaufsgruppen.

Tabelle A.4: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 4

	Health A t0	Alter	BD	Dosis PT	Anteil VT	AU-Tage	HAQ
HEALTH-49 A t0	1.00						
Alter	.059	1.00					
BD	.117**	-.014	1.00				
Dosis PT	.076	-.177***	.378***	1.00			
Anteil VT	-.089*	.024	.017	.277***	1.00		
AU-Tage	.281***	.094*	.089*	-.028	-.056	1.00	
HAQ	-.139**	.002	-.033	-.027	.039	-.007	1.00
SWE	-.523***	.152***	-.154***	-.107*	-.004	-.209***	.155***
SF-8 Psych. t0	-.623***	.089*	-.183***	-.087	.080	-.235***	.129***
FPTM LD	.596***	-.010	.231***	.100*	-.097*	.228***	-.125**
FPTM HO	-.261***	.015	-.087*	.045	.063	-.147**	.294***
FPTM VH	.164**	-.019	-.033	-.069	.017	.035	-.114**
FPTM SZ	-.072	-.114**	-.081	-.008	.008	-.099	.160**
Explorative VG	.494***	.004	.019*	.019*	.003	.072***	.043***
Theoriegel. VG	.215***	.006	.112***	.039***	.008	.002	.009
HEALTH-49 A t2	.636***	.073	.052	.002	-.050	.283***	-.249***
SF-8 Psych. t2	-.459***	.026	-.101*	-.008	-.057	-.205***	.173***

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers V; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

BD: Behandlungsdauer; Dosis Psychotherapie: Dosis psychotherapienaher Maßnahmen; Anteil VT: Anteilige Dosis an verhaltenstherapeutischen Maßnahmen; AU-Tage: Arbeitsunfähigkeitstage; FPTM LD: FPTM psychischer Leidensdruck; FPTM HO: FPTM Hoffnung; FPTM VH: FPTM Verneinung psychischer Hilfsbedürftigkeit; FPTM SZ: FPTM Symptombezogene Zuwendung durch Andere; Explorative VG: Explorative Verlaufgruppen; Theoriegel. VG: Theoriegeleitete Verlaufgruppen.

Tabelle A.5: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 5

	SWE	SF-8 t0	FPTM LD	FPTM HO	FPTM VH	FPTM SZ
SWE	1.00					
SF-8 Psych. t0	.479***	1.00				
FPTM LD	-.569***	-.656***	1.00			
FPTM HO	.240***	.245***	-.207***	1.00		
FPTM VH	-.037	-.037	-.004	-.293***	1.00	
FPTM SZ	.076	.068	-.043	.040	-.014	1.00
Explorative VG	.102***	.148***	.134***	.065***	.026**	.026**
Theoriegel. VG	.065***	.135***	.087***	.005	.011	.009
HEALTH-49 A t2	-.371***	-.346***	.403***	-.295***	.099*	-.141*
SF-8 Psych. t2	.328***	.414***	-.401**	.211***	-.093	.066

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers V; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

FPTM LD: FPTM *psychischer Leidensdruck*; FPTM HO: FPTM *Hoffnung*; FPTM VH: FPTM *Verneinung psychischer Hilfsbedürftigkeit*; FPTM SZ: FPTM *Symptombezogene Zuwendung durch Andere*; Explorative VG: Explorative Verlaufgruppen; Theoriegel. VG: Theoriegeleitete Verlaufgruppen.

Tabelle A.6: Prüfung auf Multikollinearität: Zusammenhangsmaße der nominal- und intervallskalierten Merkmale – Teil 6

	Explorative Verlaufgruppen	Theoriegeleitete Verlaufgruppen	Health-49 A t2	SF-8 Psych. t2
Explorative VG	1.00			
Theoriegel. VG	.221***	1.00		
HEALTH-49 A t2	.347***	.018	1.00	
SF-8 Psych. t2	.116***	.027*	-.714***	1.00

Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**) und dem 0.1%- Niveau (***) festgelegt.

Anmerkung: Zusammenhangsmaß für nominal-zweistufige Variablen: Phi; Zusammenhangsmaß für nominal-mehrstufige Variablen: Cramers V; Zusammenhangsmaß für nominal-zweistufige und intervallskalierte Variablen: punktbiseriale Korrelation; Zusammenhangsmaß für nominal-mehrstufige und intervallskalierte Variablen: Eta (intervallskalierte Variable als abhängige Variable).

Explorative VG: Explorative Verlaufgruppen; Theoriegel. VG: Theoriegeleitete Verlaufgruppen.

Anhang B: Stichprobenbeschreibungen und Drop-Out-Analysen

Tabelle B.1: Soziodemographische Patientenmerkmale – Vergleiche der Verlaufsstichprobe mit der Drop-Out-Stichprobe

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Geschlecht				
Frauen	490 (76.8)	447 (77.6)	43 (69.4)	$\chi^2 (1, n = 638) = 2.138$ $p = .144, \phi = -.058$
Männer	148 (23.2)	129 (22.4)	19 (30.6)	
Alter				
in Jahren (Range: 18 – 77)	43.7 (11.3)	43.9 (11.1)	41.8 (12.9)	$t(635) = -1.382$ $p = .168, d = 0.18$
Partnerschaft				
in Partnerschaft	361 (56.4)	328 (56.9)	33 (51.6)	$\chi^2 (1, n = 640) = 0.678$ $p = .410, \phi = .033$
nicht in Partnerschaft	279 (43.6)	248 (43.1)	31 (48.4)	
Schulabschluss				
hoch (ab Realschule)	467 (73.7)	421 (73.3)	46 (73.0)	$\chi^2 (1, n = 637) = 0.003$ $p = .995, \phi = .002$
niedrig (Hauptschule/ kein Schulabschluss)	170 (26.7)	153 (26.7)	17 (27.0)	
Berufsabschluss				
hoch (Meister, FH, Uni)	182 (31.1)	166 (31.3)	16 (29.1)	$\chi^2 (1, n = 586) = 0.110$ $p = .740, \phi = .014$
niedrig (Lehre, noch in Ausbildung, kein Berufsabschluss, sonstige)	404 (69.9)	365 (68.7)	39 (70.9)	
Erwerbssituation				
erwerbstätig	454 (71.0)	411 (71.4)	43 (68.3)	$\chi^2 (1, n = 639) = 0.265$ $p = .606, \phi = .020$
nicht erwerbstätig	185 (29.0)	165 (28.6)	20 (31.7)	
Familienstand /Partnerschaft				
mit Partner (verheiratet)	227 (44.5)	254 (45.0)	23 (39.7)	$\chi^2 (3, n = 622) = 1.579$ $p = .664, \phi = .050$
mit Partner (ledig)	82 (13.2)	72 (12.8)	10 (17.2)	
ohne Partner (ledig)	89 (14.3)	82 (14.5)	7 (12.1)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	174 (28.0)	156 (27.7)	18 (31.0)	

	Gesamt (N = 640)	Verlaufs- Stichprobe (n = 576)	Drop-Out- Stichprobe (n = 64)	Signifikanztest
soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2-Test / FE-Test
Schulabschluss				
Hauptschule	148 (23.7)	133 (23.7)	15 (23.8)	
Realschule/ Polytechnische Oberschule	262 (41.9)	231 (41.1)	31 (49.2)	$\chi^2 (3, n = 625) = 2.046$ $p = .563, \phi = .057$
Abitur/ Fachabitur	193 (30.9)	178 (31.7)	15 (23.8)	
sonstige oder ohne Schulabschluss	22 (3.5)	20 (3.6)	2 (3.2)	
Berufsabschluss				
noch in Ausbildung	13 (2.2)	11 (2.1)	2 (3.6)	
Lehre/Fachschule	328 (56.0)	299 (56.3)	29 (52.7)	
Meister/Techniker	81 (13.8)	70 (13.2)	11 (20.0)	FE (n = 586) = 5.972 $p = .309, \phi = .101$
Fachhochschule/ Universität	101 (17.2)	96 (18.1)	5 (9.1)	
ohne Berufsabschluss	50 (8.5)	43 (8.1)	7 (12.7)	
sonstiges	13 (2.2)	12 (2.3)	1 (1.8)	
Erwerbstätigkeit				
Vollzeit berufstätig	302 (52.4)	274 (52.6)	28 (50.9)	
Teilzeit berufstätig	142 (24.7)	128 (24.6)	15 (25.5)	$\chi^2 (2, n = 576) = 0.056$ $p = .972, \phi = .010$
nicht berufstätig	132 (22.9)	119 (22.8)	13 (23.6)	
monatliches Haushaltsnettoeinkommen				
bis 1 000 Euro	95 (14.8)	81 (14.1)	14 (21.9)	
bis 2 000 Euro	210 (32.8)	189 (32.8)	21 (32.8)	
bis 3 000 Euro	160 (25.0)	143 (24.8)	17 (26.6)	$\chi^2 (4, n = 640) = 5.549$ $p = .235, \phi = .093$
bis 4 000 Euro	75 (11.7)	72 (12.5)	3 (4.7)	
mehr als 4 000 Euro	100 (15.6)	91 (15.8)	9 (14.1)	

M: Mittelwert; SD: Standardabweichung; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; d: Effektstärke Cohens d. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (**), und große (***) Effektstärken interpretiert.

Tabelle B.2: Soziodemographische Patientenmerkmale – Vergleiche der Verlaufs-Outcome-Stichprobe mit der Katamnese-Drop-Out-Stichprobe

Soziodemographische Merkmale	Verlaufs-Stichprobe (N = 576)	Verlaufs-Outcome-Stichprobe (n = 424)	Katamnese-Drop-Out-Stichprobe (n = 152)	Signifikanztest
	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	t-Test / χ^2 -Test / FE-Test
Geschlecht				
Frauen	447 (77.6)	333 (78.5)	114 (75.0)	$\chi^2 (1, n = 576) = 0.806$ $p = .369, \phi = -.037$
Männer	129 (22.4)	91 (21.5)	38 (25.0)	
Alter				
in Jahren (Range: 18 – 77)	43.9 (11.1)	45.3 (10.5)	40.4 (12.0)	$t(574) = -4.651$ $p < .001^{***}, d = 0.47^{**}$
Partnerschaft				
in Partnerschaft	328 (56.9)	243 (57.3)	85 (55.9)	$\chi^2 (1, n = 576) = 0.088$ $p = .766, \phi = .012$
nicht in Partnerschaft	248 (43.1)	181 (42.7)	67 (44.1)	
Schulabschluss				
hoch (ab Realschule)	421 (73.3)	316 (74.9)	105 (69.1)	$\chi^2 (1, n = 574) = 1.925$ $p = .165, \phi = .058$
niedrig (Hauptschule/ kein Schulabschluss)	153 (26.7)	106 (25.1)	47 (30.9)	
Berufsabschluss				
hoch (Meister, FH, Uni)	166 (31.3)	128 (32.6)	38 (27.5)	$\chi^2 (1, n = 531) = 1.204$ $p = .272, \phi = .048$
niedrig (Lehre, noch in Ausbildung, kein Berufsabschluss, sonstige)	365 (68.7)	265 (67.4)	100 (72.5)	
Erwerbssituation				
erwerbstätig	411 (71.4)	309 (72.9)	102 (67.1)	$\chi^2 (1, n = 576) = 1.824$ $p = .177, \phi = .056$
nicht erwerbstätig	165 (28.6)	115 (27.1)	50 (32.9)	
Familienstand /Partnerschaft				
mit Partner (verheiratet)	254 (45.0)	193 (46.5)	61 (40.9)	$\chi^2 (3, n = 564) = 2.916$ $p = .405, \phi = .072$
mit Partner (ledig)	72 (12.8)	48 (11.6)	24 (16.1)	
ohne Partner (ledig)	82 (14.5)	58 (14.0)	24 (16.1)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	156 (27.7)	116 (28.0)	40 (26.8)	
Schulabschluss				
Hauptschule	133 (23.7)	94 (22.9)	39 (25.8)	$\chi^2 (3, n = 562) = 2.854$ $p = .415, \phi = .071$
Realschule/ Polytechnische Oberschule	231 (41.1)	170 (41.4)	61 (40.4)	
Abitur/ Fachabitur	178 (31.7)	135 (32.8)	43 (28.5)	
sonstige oder ohne Schulabschluss	20 (3.6)	12 (2.9)	8 (5.3)	
Berufsabschluss				

	Verlaufs- Stichprobe (N = 576)	Verlaufs- Outcome- Stichprobe (n = 424)	Katamnese- Drop-Out- Stichprobe (n = 152)	Signifikanztest
Soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>t-Test / χ^2-Test / FE-Test</i>
noch in Ausbildung	11 (2.1)	8 (2.0)	3 (2.2)	
Lehre/Fachschule	299 (56.3)	222 (56.5)	77 (55.8)	
Meister/Techniker	70 (13.2)	52 (13.2)	18 (13.0)	<i>FE (n = 531) = 4.682 p = .455, ϕ = .092</i>
Fachhochschule/ Universität	96 (18.1)	76 (19.3)	20 (14.5)	
ohne Berufsabschluss	43 (8.1)	27 (6.9)	16 (11.6)	
sonstiges	12 (2.3)	8 (2.0)	4 (2.9)	
Erwerbstätigkeit				
Vollzeit berufstätig	274 (52.6)	209 (54.7)	65 (46.8)	<i>χ^2 (2, n = 521) = 3.512 p = .173, ϕ = .082</i>
Teilzeit berufstätig	128 (24.6)	93 (24.3)	35 (25.2)	
nicht berufstätig	119 (22.8)	35 (25.2)	39 (28.1)	
monatliches Haushaltsnettoeinkommen				
bis 1 000 Euro	81 (14.1)	59 (13.9)	22 (14.5)	<i>χ^2 (4, n = 576) = 5.194 p = .268, ϕ = .095</i>
bis 2 000 Euro	189 (32.8)	130 (30.7)	59 (38.8)	
bis 3 000 Euro	143 (24.8)	108 (25.5)	35 (23.0)	
bis 4 000 Euro	72 (12.5)	59 (13.9)	13 (8.6)	
mehr als 4 000 Euro	91 (15.8)	68 (16.0)	23 (15.1)	

M: Mittelwert; *SD*: Standardabweichung; χ^2 : Chi²-Test; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi; *d*: Effektstärke Cohens *d*. Die Signifikanzgrenzen wurden auf dem 5%- (*), dem 1%- (**), und dem 0.1%-Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Anhang C: Vergleiche der explorativen Verlaufsgruppen

Tabelle C.1: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in soziodemographischen Patientenmerkmalen

	Gesamte Stichprobe (N = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / FE-Test
Geschlecht							
Frauen	447 (77.6%)	310 (75.4%)	50 (82.0%)	48 (90.6%)	21 (72.4%)	18 (81.8%)	$FE (n = 576) = 7.938$ $p = .092^{\circ}$, $\phi = -.106$
Männer	129 (22.4%)	101 (24.6%)	11 (18.0%)	5 (9.4%)	8 (27.6%)	4 (18.2%)	
Alter							
in Jahren (Range: 18 – 77 Jahre)	43.94 (11.13)	43.74 (10.91)	44.61 (11.83)	43.87 (11.61)	46.48 (11.95)	42.55 (11.37)	$F(4,571) = -0.550$ $p = .699$, $\eta^2 = .004$
≤ 29 Jahre alt	70 (12.6%)	54 (13.6%)	7 (11.9%)	4 (7.8%)	3 (10.3%)	2 (9.5%)	
30 bis 39 Jahre alt	101 (18.2%)	65 (16.4%)	9 (15.3%)	14 (27.5%)	5 (17.2%)	8 (38.1%)	$FE (n = 556) = 18.591$ $p = .253$, $\phi = .190^{+}$
40 bis 49 Jahre alt	205 (36.9%)	155 (39.1%)	19 (32.2%)	19 (37.3%)	7 (24.1%)	5 (23.8%)	
50 bis 59 Jahre alt	147 (26.4%)	101 (25.5%)	21 (35.6%)	10 (19.6%)	10 (34.5%)	5 (23.8%)	
≥ 60 Jahre alt	33 (5.9%)	21 (5.3%)	3 (5.1%)	4 (7.8%)	4 (13.8%)	1 (4.8%)	
Familienstand							
mit Partner (verheiratet)	254 (45.0%)	185 (45.8%)	24 (40.7%)	24 (47.1%)	10 (34.5%)	11 (52.4%)	$FE (n = 564) = 8.753$ $p = .724$, $\phi = .127^{+}$
mit Partner (ledig)	72 (12.8%)	50 (12.4%)	10 (16.9%)	6 (11.8%)	5 (17.2%)	1 (4.8%)	
ohne Partner (ledig)	82 (14.5%)	59 (14.6%)	9 (15.3%)	10 (19.6%)	2 (6.9%)	2 (9.5%)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	156 (27.7%)	110 (27.2%)	16 (27.1%)	11 (21.6%)	12 (41.4%)	7 (33.3%)	

	Gesamte Stichprobe (N = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / FE-Test
Schulabschluss							
Hauptschule	133 (23.7%)	79 (19.8%)	23 (37.7%)	11 (21.6%)	12 (42.9%)	8 (36.4%)	
Realschule/ Polytechnische Oberschule	231 (41.1%)	161 (40.3%)	19 (31.1%)	27 (52.9%)	13 (46.4%)	11 (50.0%)	<i>FE (n = 562) = 39.848 p < .001***, $\phi = .267^+$</i>
Abitur/ Fachabitur	178 (31.7%)	148 (37.0%)	14 (23.0%)	12 (23.5%)	3 (10.7%)	1 (4.5%)	
sonstige oder kein Schulabschluss	20 (3.6%)	12 (3.0%)	5 (8.2%)	1 (2.0%)	0	2 (9.1%)	
Berufsabschluss							
noch in der Berufsausbildung	11 (2.1%)	7 (1.9%)	3 (5.2%)	0	1 (4.0%)	0	
Lehre/Fachschule	299 (56.3%)	202 (53.4%)	34 (58.6%)	35 (68.6%)	15 (60.0%)	13 (68.4%)	
Meister/ Techniker	70 (13.2%)	53 (14.0%)	7 (12.1%)	7 (13.7%)	2 (8.0%)	1 (5.3%)	<i>FE (n = 531) = 34.999 p = .008**, $\phi = .255^+$</i>
Fachhochschule/ Universität	96 (18.1%)	85 (22.5%)	3 (5.2%)	4 (7.8%)	3 (12.0%)	1 (5.3%)	
ohne Berufsabschluss	43 (8.1%)	23 (6.1%)	10 (17.2%)	4 (7.8%)	3 (12.0%)	3 (15.8%)	
andere	12 (2.3%)	8 (2.1%)	1 (1.7%)	1 (2.0%)	1 (4.0%)	1 (5.3%)	
Erwerbstätigkeit							
Vollzeit berufstätig	274 (52.6%)	216 (57.4%)	19 (35.8%)	24 (51.1%)	9 (34.6%)	6 (31.6%)	
Teilzeit berufstätig oder geringfügig beschäftigt	128 (24.6%)	91 (24.2%)	10 (18.9%)	13 (27.7%)	8 (30.8%)	6 (31.6%)	
nicht berufstätig: in Rente	25 (4.8%)	11 (2.9%)	6 (11.3%)	2 (4.3%)	4 (15.4%)	2 (10.5%)	<i>FE (n = 521) = 34.882 p = .002**, $\phi = .263^+$</i>
nicht berufstätig: arbeitslos	61 (11.7%)	35 (9.3%)	13 (24.5%)	6 (12.8%)	4 (15.4%)	3 (15.8%)	
nicht berufstätig: sonstige (freigestellt, in Ausbildung, Schüler/Student, Umschulung, Hausfrau/-mann)	33 (6.3%)	23 (6.1%)	5 (9.4%)	2 (4.3%)	1 (3.8%)	2 (10.5%)	

	Gesamte Stichprobe (N = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / FE-Test
Erwerbstätigkeit							
Vollzeit berufstätig	274 (52.6%)	216 (57.4%)	19 (35.8%)	24 (51.1%)	9 (34.6%)	6 (31.6%)	<i>FE (n = 521) = 26.151 p = .001**, $\phi = .230^+$</i>
Teilzeit berufstätig	128 (24.6%)	91 (24.2%)	10 (18.9%)	13 (27.7%)	8 (30.8%)	6 (31.6%)	
nicht berufstätig	119 (22.8%)	69 (18.4%)	24 (45.3%)	10 (21.3%)	9 (34.6%)	7 (36.8%)	
Monatliches Haushaltsnettoeinkommen							
bis 1 000 Euro	81 (14.1%)	48 (11.7%)	16 (26.2%)	7 (13.2%)	7 (24.1%)	3 (13.6%)	<i>FE (n = 576) = 21.546 p = .137, $\phi = .199^+$</i>
bis 2 000 Euro	189 (32.8%)	135 (32.8%)	18 (29.5%)	21 (39.6%)	6 (20.7%)	9 (40.9%)	
bis 3 000 Euro	143 (24.8%)	104 (25.3%)	13 (21.3%)	14 (26.4%)	5 (17.2%)	7 (31.8%)	
bis 4 000 Euro	72 (12.5%)	56 (13.6%)	3 (4.9%)	6 (11.3%)	6 (20.7%)	1 (4.5%)	
mehr als 4 000 Euro	91 (15.8%)	68 (16.5%)	11 (18.0%)	5 (9.4%)	5 (17.2%)	2 (9.1%)	

M: Mittelwert; *SD*: Standardabweichung; ANOVA: Varianzanalyse; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (**), und große (***) Effektstärken interpretiert.

Tabelle C.2: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in sozialmedizinischen Patientenmerkmalen

	Gesamte Stichprobe (N = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
sozialmedizinische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / χ^2 -Test / FE-Test
Arbeitsunfähigkeits-/Krankheitszeiten in den letzten 6 Monaten vor Aufnahme							
in Wochen (Range 0 – 26)	9.83 (9.40%)	8.39 (8.93%)	15.15 (9.38%)	10.46 (9.61%)	14.84 (8.78%)	13.66 (10.33%)	$F(4,533) = 10.366$ $p < .001^{***}$, $\eta^2 = .072^{**}$
Stationäre psychotherapeutische Vorbehandlung							
psychotherapeutische Krankenhausbehandlung in den 6 Monaten vor Aufnahme	58 (10.1%)	34 (8.3%)	11 (18.0%)	4 (7.5%)	6 (20.7%)	3 (13.6%)	$FE (n = 574) = 9.610$ $p = .036^*$, $\phi = .132^+$
keine	516 (89.9%)	375 (91.7%)	50 (82.0%)	49 (92.5%)	23 (79.3%)	19 (86.4%)	
Ambulante psychotherapeutische Vorbehandlung							
Einzelpsychotherapie in den 6 Monaten vor Aufnahme	277 (48.5%)	185 (45.5%)	38 (62.3%)	26 (49.1%)	15 (53.6%)	13 (59.1%)	$\chi^2 (4, n = 571) = 7.442$ $p = .114$, $\phi = .114^+$
keine	294 (51.5%)	222 (54.5%)	23 (37.7%)	27 (50.9%)	13 (46.4%)	9 (40.9%)	

M: Mittelwert; SD: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (**), und große (***) Effektstärken interpretiert.

Tabelle C.3: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in Prozessmerkmalen

	Gesamte Stichprobe (n = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
Prozessmerkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / χ^2 -Test / FE-Test
Behandlungsdauer							
in Tagen (Range 19 - 60)	41.98 (8.51)	41.63 (8.45)	44.70 (8.53)	41.72 (9.02)	43.90 (7.60)	39.00 (8.21)	$F(4,571) = 2.827$ $p = .024^*$, $\eta^2 = .019^+$
3 - 4 Wochen	74 (12.8%)	55 (13.4%)	5 (8.2%)	8 (15.1%)	2 (6.9%)	4 (18.2%)	
5 Wochen	74 (12.8%)	60 (14.6%)	3 (4.9%)	5 (9.4%)	2 (6.9%)	4 (18.2%)	
6 Wochen	290 (50.3%)	203 (49.4%)	31 (50.8%)	29 (54.7%)	16 (55.2%)	11 (50.0%)	$FE(n = 576) = 17.262$ $p = .332$, $\phi = .176^+$
7 Wochen	55 (9.5%)	38 (9.2%)	7 (11.5%)	3 (5.7%)	5 (17.2%)	2 (9.1%)	
8 - 9 Wochen	83 (14.4%)	55 (13.4%)	15 (24.6%)	8 (15.1%)	4 (13.8%)	1 (4.5%)	
Dosis psychotherapeutischer Maßnahmen							
Gesamtdauer (in Minuten)	2 060 (845)	2 008 (791)	2 149 (975)	2 313 (873)	2 275 (1 083)	1 838 (865)	$F(4,510) = 2.415$ $p = .048^*$, $\eta^2 = .019^+$
Dosis therapieschulenspezifischer Maßnahmen							
Anteil verhaltensther. zu psychodynamischen Maßnahmen, %	33.2 (30.9)	34.0 (32.2)	31.9 (28.4)	33.4 (25.5)	26.7 (25.3)	31.3 (32.2)	$F(4,508) = 0.411$ $p = .801$, $\eta^2 = .003$
Klinik							
Klinik 1	186 (32.3%)	123 (29.9%)	23 (37.7%)	23 (43.4%)	10 (34.5%)	7 (31.8%)	
Klinik 2	237 (41.1%)	189 (46.0%)	19 (31.1%)	13 (24.5%)	9 (31.0%)	7 (31.8%)	$\chi^2(8, n = 576) = 15.416$ $p = .052^{\circ}$, $\phi = .164^+$
Klinik 3	153 (26.6%)	99 (24.1%)	19 (31.1%)	17 (32.1%)	10 (34.5%)	8 (36.4%)	

M: Mittelwert; SD: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- ([°]), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Tabelle C.4: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in klinischen Patientenmerkmalen

		Gesamte Stichprobe (n = 571)	Linear Response (n = 407)	Non-Response (n = 60)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
klinische Merkmale		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	χ^2 -Test / FE-Test
Diagnose								
Unipolare Depression	F32, F33, F34.1	385 (67.4%)	264 (64.9%)	41 (68.3%)	42 (79.2%)	24 (82.8%)	14 (63.6%)	χ^2 (4, n = 571) = 7.857 $p = .097^\circ$, $\phi = .117^+$
Angststörung	F40, F41, F48	105 (18.4%)	62 (15.2%)	19 (31.7%)	13 (24.5%)	8 (27.6%)	3 (13.6%)	FE (n = 571) = 12.557 $p = .011^*$, $\phi = .151^+$
PTBS	F43.1	31 (5.4%)	15 (3.7%)	4 (6.7%)	6 (11.3%)	4 (13.8%)	2 (9.1%)	FE (n = 571) = 10.964 $p = .013^*$, $\phi = .137^+$
Anpassungsstörung	F43.2	103 (18.0%)	81 (19.9%)	13 (21.7%)	5 (9.4%)	2 (6.9%)	2 (9.1%)	FE (n = 571) = 7.446 $p = .106$, $\phi = .117^+$
Somatoforme Störung	F45	81 (14.2%)	45 (11.1%)	11 (18.3%)	16 (30.2%)	7 (24.1%)	2 (9.1%)	FE (n = 571) = 16.430 $p = .002^{**}$, $\phi = .178^+$
Essstörung	F50	96 (16.8%)	62 (15.2%)	13 (21.7%)	8 (15.1%)	3 (10.3%)	10 (45.5%)	FE (n = 571) = 12.848 $p = .009^{**}$, $\phi = .165^+$
Persönlichkeitsstörung	F60 – F61	36 (6.3%)	18 (4.4%)	6 (10.0%)	2 (3.8%)	6 (20.7%)	4 (18.2%)	FE (n = 571) = 16.483 $p = .001^{***}$, $\phi = .186^+$
Adipositas	E66	151 (26.4%)	95 (23.3%)	20 (33.3%)	21 (39.6%)	8 (27.6%)	7 (31.8%)	FE (n = 571) = 8.556 $p = .073^\circ$, $\phi = .122^+$
andere psych. Störung		79 (13.8)	54 (13.3%)	5 (8.3%)	6 (11.3%)	7 (24.1%)	7 (31.8%)	FE (n = 571) = 9.464 $p = .042^*$, $\phi = .135^+$
Anzahl psychischer Störungen (Anzahl F-Diagnosen inkl. E66)								
1 F-Diagnose		254 (44.5%)	215 (52.8%)	12 (20.0%)	12 (22.6%)	6 (20.7%)	9 (40.9%)	
2 F-Diagnosen		204 (35.7%)	129 (31.7%)	34 (56.7%)	25 (47.2%)	10 (34.5%)	6 (27.3%)	
3 F-Diagnosen		87 (15.2%)	53 (13.0%)	8 (13.3%)	13 (24.5%)	9 (31.0%)	4 (18.2%)	FE (n = 571) = 71.342 $p < .001^{***}$, $\phi = .367^{**}$
4 F-Diagnosen		22 (3.8%)	9 (2.2%)	6 (10.0%)	1 (1.9%)	4 (13.8%)	2 (9.1%)	
5 F-Diagnosen		4 (0.7%)	1 (0.2%)	0	2 (3.8%)	0	1 (4.5%)	

	Gesamte Stichprobe (n = 571)	Linear Response (n = 407)	Non-Response (n = 60)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
klinische Merkmale	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	χ^2 -Test / FE-Test
Psychische Komorbidität							
mehr als eine F-Diagnose	317 (55.5%)	192 (47.2%)	48 (80.0%)	41 (77.4%)	23 (79.3%)	13 (59.1%)	$\chi^2 (4, n = 571) = 43.033$ $p < .001^{***}, \phi = .275^+$

χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Tabelle C.5: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in personenbezogenen Merkmalen

	Gesamte Stichprobe (n = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
personenbezogene Merkmale (bei Aufnahme)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / FE-Test
Symptombelastung							
HEALTH-49 A (Range 0 – 3.72)	1.41 (0.78)	1.07 (0.54)	2.01 (0.63)	2.29 (0.48)	2.82 (0.61)	1.96 (0.74)	$F(4,571) = 139.573$ $p < .001^{***}, \eta^2 = .494^{+++}$
HEALTH-49 A: unterhalb des Cut-Off-Wertes (0.685)	112 (19.4%)	110 (26.0%)	0	0	0	2 (9.1%)	$FE (n = 576) = 64.846$ $p < .001^{***}, \phi = .295^+$
Psychotherapiemotivation							
FPTM-Skala <i>Psychischer Leidensdruck</i> (Range 1-4)	3.17 (0.73)	3.00 (0.76)	3.56 (0.41)	3.52 (0.45)	3.78 (0.34)	3.47 (0.49)	$F(4,567) = 21.983$ $p < .001^{***}, \eta^2 = .134^{++}$
FPTM-Skala <i>Hoffnung</i> (Range 1-4)	3.17 (0.67)	3.27 (0.64)	2.91 (0.69)	3.06 (0.69)	2.71 (0.78)	2.87 (0.52)	$F(4,564) = 9.846$ $p < .001^{***}, \eta^2 = .065^{++}$
FPTM-Skala <i>Verneinung psych. Hilfsbedürftigkeit</i> (Range 1-4)	1.59 (0.66)	1.52 (0.61)	1.79 (0.75)	1.76 (0.74)	1.76 (0.94)	1.57 (0.63)	$F(4,567) = 3.856$ $p = .004^{**}, \eta^2 = .026^+$

	Gesamte Stichprobe (n = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
personenbezogene Merkmale (bei Aufnahme)	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / FE-Test
Psychotherapiemotivation (Fortsetzung)							
FPTM-Skala <i>Wissen</i> (Range 1-4)	2.66 (0.93)	2.63 (0.96)	2.60 (0.96)	2.76 (0.79)	3.04 (0.83)	2.81 (0.95)	$F(4,565) = 1.707$ $p = .147, \eta^2 = .012^+$
FPTM-Skala <i>Initiative</i> (Range 1-4)	2.96 (0.97)	3.01 (0.96)	2.74 (1.02)	2.86 (1.01)	2.66 (0.89)	3.11 (0.97)	$F(4,564) = 2.011$ $p = .091^\circ, \eta^2 = .014^+$
FPTM-Skala <i>Symptom- bezogene Zuwendung</i> (Range 1-4)	2.17 (0.77)	2.22 (0.77)	1.90 (0.67)	2.16 (0.77)	1.87 (0.74)	2.29 (0.84)	$F(4,564) = 3.789$ $p = .005^{**}, \eta^2 = .026^+$
Allgemeine Selbstwirksamkeitserwartung							
SWE (Range 10-40)	23.48 (6.51)	24.75 (6.45)	21.28 (5.75)	20.31 (5.26)	18.95 (5.15)	19.21 (5.59)	$F(4,567) = 16.034$ $p < .001^{***}, \eta^2 = .102^{++}$
Psychische Aspekte der gesundheitsbezogenen Lebensqualität							
SF-8 Psych. (Range 7-62)	32.08 (12.16)	35.00 (12.16)	25.55 (8.78)	23.92 (7.50)	22.97 (6.98)	27.36 (12.71)	$F(4,560) = 24.275$ $p < .001^{***}, \eta^2 = .148^{+++}$

M: Mittelwert; *SD*: Standardabweichung; ANOVA: Varianzanalyse; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- ([°]), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (⁺⁺) und große (⁺⁺⁺) Effektstärken interpretiert.

Tabelle C.6: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in personenbezogenen Prozessmerkmalen

	Gesamte Stichprobe (n = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
personenbezogene Prozessmerkmale	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	ANOVA
Therapeutische Beziehung							
HAQ 1 nach Woche 2-3 (Range 6 - 36)	29.12 (4.92)	29.58 (4.52)	26.96 (5.69)	29.69 (4.95)	26.55 (5.15)	19.21 (5.59)	$F(4,547) = 6.114$ $p < .001^{***}, \eta^2 = .043^+$

M: Mittelwert; *SD*: Standardabweichung; ANOVA: Varianzanalyse; *p*: Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Tabelle C.7: Univariate Analysen: Vergleich der explorativen Verlaufsgruppen in soziodemographischen dichotomen Patientenmerkmalen

	Gesamte Stichprobe (n = 576)	Linear Response (n = 411)	Non-Response (n = 61)	Early Response (n = 53)	Delayed Response (n = 29)	Atypical Course (n = 22)	Signifikanztest
soziodemographische Merkmale	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	χ^2-Test
Partnersituation							
mit Partner (verheiratet oder ledig)	328 (56.9%)	236 (57.4%)	34 (55.7%)	31 (58.5%)	15 (51.7%)	12 (54.5%)	$\chi^2 (4, n = 576) = 0.500$ $p = .974, \phi = .029$
ohne Partner (ledig, verwitwet, geschieden, getrennt lebend)	248 (43.1%)	175 (42.6%)	27 (44.3%)	22 (41.5%)	14 (48.3%)	10 (45.5%)	
Schulabschluss							
hoch (Realschule, Abitur)	421 (73.3%)	318 (77.8%)	33 (54.1%)	41 (77.4%)	17 (58.6%)	12 (54.5%)	$\chi^2 (4, n = 574) = 23.249$ $p < .001^{***}, \phi = .201^+$
niedrig (kein, Hauptschule, andere)	153 (26.7%)	91 (22.2%)	28 (45.9%)	12 (22.6%)	12 (41.4%)	10 (45.5%)	
Berufsabschluss							
hoch (Meister, FH, Uni)	166 (31.3%)	138 (36.5%)	10 (17.2%)	11 (21.6%)	5 (20.0%)	2 (10.5%)	$\chi^2 (4, n = 531) = 17.654$ $p = .001^{***}, \phi = .182^+$
niedrig (kein, Lehre, andere)	365 (68.7%)	240 (63.5%)	48 (82.8%)	40 (78.4%)	20 (80.0%)	17 (89.5%)	
Erwerbstätigkeit							
erwerbstätig	411 (71.4%)	314 (76.4%)	29 (47.5%)	38 (71.7%)	18 (62.1%)	12 (54.5%)	$\chi^2 (4, n = 576) = 26.308$ $p < .001^{***}, \phi = .214^+$
nicht erwerbstätig	165 (28.6%)	97 (23.6%)	32 (52.5%)	15 (28.3%)	11 (37.9%)	10 (45.5%)	

χ^2 : Chi²-Test; p : Signifikanz; ϕ : Effektstärke Phi. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Anhang D: Vergleiche der theoriegeleiteten Verlaufgruppen

Tabelle D.1: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in soziodemographischen Patientenmerkmalen

	Gesamte Stichprobe (n = 576)	Non-Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
soziodemographische Merkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / χ^2 -Test / FE-Test
Geschlecht						
Frauen	447 (77.6%)	100 (71.9%)	233 (81.2%)	46 (82.1%)	68 (72.3%)	$\chi^2 (3, n = 576) = 6.843$ $p = .075^\circ, \phi = -.109^+$
Männer	129 (22.4%)	39 (28.1%)	54 (18.8%)	10 (17.9%)	26 (27.7%)	
Alter						
in Jahren (Range: 18 – 77 Jahre)	43.94 (11.13)	42.93 (10.47)	44.27 (11.31)	42.70 (12.11)	45.16 (10.87)	$F(3,572) = 1.079$ $p = 358, \eta^2 = .006$
≤ 29 Jahre alt	70 (12.6%)	16 (12.1%)	35 (12.6%)	10 (18.2%)	9 (9.8%)	$FE (n = 556) = 8.404$ $p = .756, \phi = .123^+$
30 bis 39 Jahre alt	101 (18.2%)	29 (22.0%)	46 (16.6%)	12 (21.8%)	14 (15.2%)	
40 bis 49 Jahre alt	205 (36.9%)	52 (39.4%)	100 (36.1%)	16 (29.1%)	37 (40.2%)	
50 bis 59 Jahre alt	147 (26.4%)	30 (22.7%)	78 (28.2%)	14 (25.5%)	25 (27.2%)	
≥ 60 Jahre alt	33 (5.9%)	5 (3.8%)	18 (6.5%)	3 (5.5%)	7 (7.6%)	
Familienstand						
mit Partner (verheiratet)	254 (45.0%)	62 (45.6%)	125 (44.3%)	20 (36.4%)	47 (51.6%)	$\chi^2 (9, n = 564) = 15.226$ $p = .085^\circ, \phi = .164^+$
mit Partner (ledig)	72 (12.8%)	22 (16.2%)	31 (11.0%)	10 (18.2%)	9 (9.9%)	
ohne Partner (ledig)	82 (14.5%)	26 (19.1%)	41 (14.5%)	5 (9.1%)	10 (11.0%)	
ohne Partner (verwitwet, geschieden, getrennt lebend)	156 (27.7%)	26 (19.1%)	85 (30.1%)	20 (36.4%)	25 (27.5%)	

	Gesamte Stichprobe (n = 576)	Non- Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / χ^2-Test / FE-Test
Höchster Schulabschluss						
Hauptschule	133 (23.7%)	34 (25.6%)	62 (22.2%)	13 (23.2%)	24 (25.5%)	
Realschule/ Polytechnische Oberschule	231 (41.1%)	56 (42.1%)	114 (40.9%)	24 (42.9%)	37 (39.4%)	<i>FE (n = 562) = 4.119 p = .912, ϕ = .087</i>
Abitur/ Fachabitur	178 (31.7%)	41 (30.8%)	91 (32.6%)	18 (32.1%)	28 (29.8%)	
sonstige oder kein Schulabschluss	20 (3.6%)	2 (1.5%)	12 (4.3%)	1 (1.8%)	5 (5.3%)	
Höchster Berufsabschluss						
noch in der Berufsausbildung	11 (2.1%)	4 (3.1%)	5 (1.9%)	1 (1.9%)	1 (1.3%)	
Lehre/Fachschule	299 (56.3%)	72 (55.8%)	148 (55.0%)	27 (50.9%)	52 (65.0%)	
Meister/ Techniker	70 (13.2%)	18 (14.0%)	33 (12.3%)	12 (22.6%)	7 (8.8%)	<i>FE (n = 531) = 11.906 p = .662, ϕ = .154⁺</i>
Fachhochschule/ Universität	96 (18.1%)	22 (17.1%)	51 (19.0%)	7 (13.2%)	16 (20.0%)	
ohne Berufsabschluss	43 (8.1%)	9 (7.0%)	25 (9.3%)	5 (9.4%)	4 (5.0%)	
andere	12 (2.3%)	4 (3.1%)	7 (2.6%)	1 (1.9%)	0	
Erwerbstätigkeit						
Vollzeit berufstätig	274 (52.6%)	75 (59.1%)	129 (50.0%)	31 (62.0%)	39 (45.3%)	
Teilzeit berufstätig	128 (24.6%)	33 (26.0%)	64 (24.8%)	5 (10.0%)	26 (30.2%)	<i>χ^2 (6, n = 521) = 13.344 p = .038*, ϕ = .160⁺</i>
nicht berufstätig	119 (22.8%)	19 (15.0%)	65 (25.2%)	14 (28.0%)	21 (24.4%)	

	Gesamte Stichprobe (n = 576)	Non- Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
soziodemographische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / χ^2-Test / FE-Test
Monatliches Haushaltsnettoeinkommen						
bis 1 000 Euro	81 (14.1%)	19 (13.7%)	44 (15.3%)	7 (12.5%)	11 (11.7%)	
bis 2 000 Euro	189 (32.8%)	35 (25.2%)	106 (39.6%)	22 (39.3%)	26 (27.7%)	
bis 3 000 Euro	143 (24.8%)	43 (30.9%)	63 (22.0%)	10 (17.9%)	27 (28.7%)	$\chi^2 (12, n = 576) = 21.546$ $p = .339, \phi = .153^+$
bis 4 000 Euro	72 (12.5%)	17 (12.2%)	35 (12.2%)	6 (10.7%)	14 (14.9%)	
mehr als 4 000 Euro	91 (15.8%)	25 (18.0%)	39 (13.6%)	11 (19.6%)	16 (17.0%)	

M: Mittelwert; *SD*: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Tabelle D.2: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufsgruppen in sozialmedizinischen Patientenmerkmalen

	Gesamte Stichprobe (n = 576)	Non- Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
sozialmedizinische Merkmale	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / χ^2-Test
Arbeitsunfähigkeits- / Krankheitszeiten in den letzten 6 Monaten vor Aufnahme						
in Wochen (Range 0 – 26)	9.83 (9.40%)	9.47 (9.44%)	10.04 (9.15%)	10.69 (10.44%)	9.21 (9.53%)	$F(3,534) = 0.377$ $p = .769, \eta^2 = .002$
Stationäre psychotherapeutische Vorbehandlung						
psychotherapeutische Krankenhausbehandlung in den 6 Monaten vor Aufnahme	58 (10.1%)	16 (11.6%)	31 (10.8%)	4 (7.1%)	7 (7.4%)	$\chi^2 (3, n = 574) = 1.779$ $p = .625, \phi = .056$
keine	516 (89.9%)	122 (88.4%)	255 (89.2%)	52 (92.9%)	87 (92.6%)	
Ambulante psychotherapeutische Vorbehandlung						
Einzelpsychotherapie in den 6 Monaten vor Aufnahme	277 (48.5%)	69 (50.7%)	140 (49.1%)	25 (44.6%)	43 (45.7%)	$\chi^2 (3, n = 571) = 0.936$ $p = .811, \phi = .040$
keine	294 (51.5%)	67 (49.3%)	145 (50.9%)	31 (55.4%)	51 (54.3%)	

M: Mittelwert; *SD*: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; *p*: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Tabelle D.3: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufsgruppen in Prozessmerkmalen

	Gesamte Stichprobe (n = 576)	Non- Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
Prozessmerkmale	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / χ^2-Test / FE-Test
Behandlungsdauer						
in Tagen (Range 19 – 60)	41.98 (8.51)	43.47 (6.51)	42.52 (8.03)	45.71 (7.38)	35.88 (10.23)	$F(3,572) = 24.042$ $p < .001^{***}$, $\eta^2 = .112^{++}$
3 - 4 Wochen	74 (12.8%)	3 (2.2%)	30 (10.5%)	2 (3.6%)	39 (41.5%)	
5 Wochen	74 (12.8%)	21 (15.1%)	39 (13.6%)	6 (10.7%)	8 (8.5%)	
6 Wochen	290 (50.3%)	81 (58.3%)	151 (52.6%)	24 (42.9%)	34 (36.2%)	$\chi^2 (12, n = 576) = 103.197$ $p < .001^{***}$, $\phi = .244^+$
7 Wochen	55 (9.5%)	15 (10.8%)	21 (7.3%)	12 (21.4%)	7 (7.4%)	
8 - 9 Wochen	83 (14.4%)	19 (13.7%)	46 (16.0%)	12 (21.4%)	6 (6.4%)	
Dosis psychotherapeutischer Maßnahmen						
Gesamtdauer (in Minuten)	2 060 (845)	2 122 (741)	2 109 (885)	2 242 (929)	1 674 (707)	$F(3,511) = 6.874$ $p < .001^{***}$, $\eta^2 = .039^+$
Dosis therapieschulenspezifischer Maßnahmen						
Anteil verhaltensther. zu psychodyn. Maßnahmen, %	33.2 (30.9)	36.7 (31.6)	30.6 (29.3)	35.1 (32.3)	35.0 (33.7)	$F(3,509) = 1.340$ $p = .260$, $\eta^2 = .008$
Klinik						
Klinik 1	186 (32.3%)	41 (29.5%)	110 (38.3%)	19 (33.9%)	16 (17.0%)	
Klinik 2	237 (41.1%)	62 (44.6%)	96 (33.4%)	21 (37.5%)	58 (61.7%)	$\chi^2 (6, n = 576) = 26.174$ $p < .001^{***}$, $\phi = .213^+$
Klinik 3	153 (26.6%)	36 (25.9%)	81 (28.2%)	16 (28.6%)	20 (21.3%)	

M: Mittelwert; SD: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; FE: Fishers exakter Test; p: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (°), mittlere (**) und große (***) Effektstärken interpretiert.

Tabelle D.4: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufsgruppen in klinischen Patientenmerkmalen

		Gesamte Stichprobe (n = 571)	Non- Response (n = 138)	Early Response (n = 285)	Late Response (n = 56)	n. a. (n = 92)	Signifikanztest
klinische Merkmale		n (%)	n (%)	n (%)	n (%)	n (%)	χ^2-Test / FE-Test
Diagnose							
Unipolare Depression	F32, F33, F34.1	385 (67.4%)	78 (56.5%)	210 (73.7%)	44 (78.6%)	53 (57.6%)	$\chi^2 (3, n = 571) = 19.757$ $p < .001^{***}$, $\phi = .186^+$
Angststörung	F40, F41, F48	105 (18.4%)	21 (15.2%)	52 (18.2%)	15 (26.8%)	17 (18.5%)	$\chi^2 (3, n = 571) = 3.560$ $p = .308$, $\phi = .079$
PTBS	F43.1	31 (5.4%)	7 (5.1%)	19 (6.7%)	3 (5.4%)	2 (2.2%)	FE (n = 571) = 2.661 $p = .432$, $\phi = .070$
Anpassungs- störung	F43.2	103 (18.0%)	34 (24.6%)	41 (14.4%)	6 (10.7%)	22 (23.9%)	$\chi^2 (3, n = 571) = 10.816$ $p = .013^*$, $\phi = .138^+$
Somatoforme Störung	F45	81 (14.2%)	19 (13.8%)	43 (15.1%)	6 (10.7%)	13 (14.1%)	$\chi^2 (3, n = 571) = 0.765$ $p = .858$, $\phi = .037$
Essstörung	F50	96 (16.8%)	25 (18.1%)	55 (19.3%)	6 (10.7%)	10 (10.9%)	$\chi^2 (3, n = 571) = 5.239$ $p = .150$, $\phi = .096$
Persönlichkeits- störung	F60 – F61	36 (6.3%)	10 (7.2%)	14 (4.9%)	5 (8.9%)	7 (7.6%)	FE (n = 571) = 2.507 $p = .471$, $\phi = .060$
Adipositas	E66	151 (26.4%)	34 (24.6%)	79 (27.7%)	17 (30.4%)	21 (22.8%)	$\chi^2 (3, n = 571) = 1.530$ $p = .683$, $\phi = .052$
andere psych. Störung		79 (13.8)	13 (9.4%)	46 (16.1%)	5 (8.9%)	15 (16.3%)	$\chi^2 (3, n = 571) = 5.128$ $p = .163$, $\phi = .095$
Anzahl psychischer Störungen (Anzahl F-Diagnosen inkl. E66)							
1 F-Diagnose		254 (44.5%)	68 (49.3%)	113 (39.6%)	23 (41.1%)	50 (54.3%)	
2 F-Diagnosen		204 (35.7%)	52 (37.7%)	105 (36.8%)	21 (37.5%)	26 (28.3%)	
3 F-Diagnosen		87 (15.2%)	14 (10.1%)	53 (18.6%)	8 (14.3%)	12 (13.0%)	FE (n = 571) = 14.389 $p = .217$, $\phi = .160^+$
4 F-Diagnosen		22 (3.8%)	4 (2.9%)	11 (3.9%)	3 (5.4%)	4 (4.3%)	
5 F-Diagnosen		4 (0.7%)	0	3 (1.1%)	1 (1.8%)	0	

	Gesamte Stichprobe (n = 571)	Non-Response (n = 138)	Early Response (n = 285)	Late Response (n = 56)	n. a. (n = 92)	Signifikanztest
klinische Merkmale	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	χ^2 -Test / <i>FE</i> -Test
Psychische Komorbidität						
mehr als eine F-Diagnose	317 (55.5%)	70 (50.7%)	172 (60.4%)	33 (58.9%)	42 (45.7%)	$\chi^2 (3, n = 571) = 7.869$ $p = .049^*$, $\phi = .117^+$

χ^2 : Chi²-Test; *FE*: Fishers exakter Test; *p*: Signifikanz; ϕ : Effektstärke Phi. Die Signifikanzgrenzen wurden auf dem 10%- ([°]), dem 5%- (^{*}), dem 1%- (^{**}) und dem 0.1%- Niveau (^{***}) festgelegt. Die Effektstärkenmaße werden als kleine (⁺), mittlere (^{**}) und große (⁺⁺⁺) Effektstärken interpretiert.

Tabelle D.5: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufsgruppen in personenbezogenen Merkmalen

	Gesamte Stichprobe (n = 571)	Non-Response (n = 138)	Early Response (n = 285)	Late Response (n = 56)	n. a. (n = 92)	Signifikanztest
personenbezogene Merkmale (bei Aufnahme)	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	<i>n (%) / M (SD)</i>	ANOVA / χ^2 -Test
Symptombelastung						
HEALTH-49 A (Range 0 – 3.72)	1.41 (0.78)	0.87 (0.63)	1.73 (0.70)	1.48 (0.68)	1.17 (0.76)	$F(3,572) = 52.210$ $p < .001^{***}$, $\eta^2 = .215^{+++}$
HEALTH-49 A: unterhalb des Cut-Off-Wertes (0.685)	112 (19.4%)	70 (50.4%)	9 (3.1%)	5 (8.9%)	28 (29.8%)	$\chi^2 (3, n = 576) = 143.921$ $p < .001^{***}$, $\phi = .500^{+++}$
Psychotherapiemotivation						
FPTM-Skala <i>Psychischer Leidensdruck</i> (Range 1-4)	3.17 (0.73)	2.88 (0.82)	3.34 (0.59)	3.38 (0.64)	2.96 (0.83)	$F(3,568) = 18.082$ $p < .001^{***}$, $\eta^2 = .087^{**}$
FPTM-Skala <i>Hoffnung</i> (Range 1-4)	3.17 (0.67)	3.25 (0.70)	3.14 (0.65)	3.17 (0.70)	3.15 (0.71)	$F(3,565) = 0.859$ $p = .462$, $\eta^2 = .005$
FPTM-Skala <i>Verneinung psych. Hilfsbedürftigkeit</i> (Range 1-4)	1.59 (0.66)	1.50 (0.57)	1.65 (0.70)	1.48 (0.60)	1.58 (0.70)	$F(3,568) = 2.182$ $p = .089^{\circ}$, $\eta^2 = .011^+$

	Gesamte Stichprobe (n = 571)	Non- Response (n = 138)	Early Response (n = 285)	Late Response (n = 56)	n. a. (n = 92)	Signifikanztest
personenbezogene Merkmale (bei Aufnahme)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	n (%) / M (SD)	ANOVA / χ^2-Test
Psychotherapiemotivation (Fortsetzung)						
FPTM-Skala <i>Wissen</i> (Range 1-4)	2.66 (0.93)	2.60 (0.95)	2.67 (0.92)	2.88 (0.90)	2.61 (0.93)	$F(3,566) = 1.297$ $p = .274, \eta^2 = .007$
FPTM-Skala <i>Initiative</i> (Range 1-4)	2.96 (0.97)	2.94 (0.99)	2.90 (0.98)	3.11 (0.91)	3.07 (0.93)	$F(3,565) = 1.260$ $p = .287, \eta^2 = .007$
FPTM-Skala <i>Symptom- bezogene Zuwendung</i> (Range 1-4)	2.17 (0.77)	2.21 (0.79)	2.21 (0.75)	2.09 (0.69)	2.03 (0.82)	$F(3,565) = 1.619$ $p = .184, \eta^2 = .009$
Allgemeine Selbstwirksamkeitserwartung						
SWE (Range 10-40)	23.48 (6.51)	25.54 (6.91)	22.32 (5.85)	21.18 (5.60)	25.32 (7.06)	$F(3,568) = 13.261$ $p < .001^{***}, \eta^2 = .065^{**}$
Psychische Aspekte der Gesundheitsbezogenen Lebensqualität						
SF-8 Psych. (Range 7-62)	32.08 (12.16)	38.34 (13.16)	28.18 (10.07)	30.20 (11.54)	36.07 (12.04)	$F(3,561) = 29.194$ $p < .001^{***}, \eta^2 = .135^{**}$

M: Mittelwert; SD: Standardabweichung; ANOVA: Varianzanalyse; χ^2 : Chi²-Test; p: Signifikanz; ϕ : Effektstärke Phi; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (**) und große (***) Effektstärken interpretiert.

Tabelle D.6: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufsgruppen in personenbezogenen Prozessmerkmalen

	Gesamte Stichprobe (n = 571)	Non- Response (n = 138)	Early Response (n = 285)	Late Response (n = 56)	n. a. (n = 92)	Signifikanztest
personenbezogene Prozessmerkmale	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	ANOVA
Therapeutische Beziehung						
HAQ 1 nach Woche 2-3 (Range 6 - 36)	29.12 (4.92)	28.62 (4.83)	29.56 (4.67)	28.93 (5.09)	28.46 (5.80)	$F(3,548) = 1.719$ $p = .162, \eta^2 = .009$

M: Mittelwert; SD: Standardabweichung; ANOVA: Varianzanalyse; p: Signifikanz; η^2 : Effektstärke Eta². Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Tabelle D.7: Univariate Analysen: Vergleich der theoriegeleiteten Verlaufgruppen in soziodemographischen dichotomen Patientenmerkmalen

	Gesamte Stichprobe (n = 576)	Non- Response (n = 139)	Early Response (n = 287)	Late Response (n = 56)	n. a. (n = 94)	Signifikanztest
soziodemographische Merkmale	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>	χ^2 -Test
Partnersituation						
mit Partner (verheiratet oder ledig)	328 (56.9%)	84 (60.4%)	157 (54.7%)	30 (53.6%)	57 (60.6%)	$\chi^2 (3, n = 576) = 2.060$ $p = .560, \phi = .060$
ohne Partner (ledig, verwitwet, geschieden, getrennt lebend)	248 (43.1%)	55 (39.6%)	130 (45.3%)	26 (46.4%)	37 (39.4%)	
Schulabschluss						
hoch (Realschule, Abitur)	421 (73.3%)	102 (73.9%)	212 (74.1%)	42 (75.0%)	65 (69.1%)	$\chi^2 (3, n = 574) = 1.037$ $p = .792, \phi = .043$
niedrig (kein, Hauptschule, andere)	153 (26.7%)	36 (26.1%)	74 (25.9%)	14 (25.0%)	29 (30.9%)	
Berufsabschluss						
hoch (Meister, FH, Uni)	166 (31.3%)	40 (31.0%)	84 (31.2%)	19 (35.8%)	23 (28.8%)	$\chi^2 (3, n = 531) = 0.758$ $p = .860, \phi = .038$
niedrig (kein, Lehre, andere)	365 (68.7%)	89 (69.0%)	185 (68.8%)	34 (64.2%)	57 (71.2%)	
Erwerbstätigkeit						
erwerbstätig	411 (71.4%)	108 (77.7%)	200 (69.7%)	38 (67.9%)	65 (69.1%)	$\chi^2 (3, n = 576) = 3.686$ $p = .297, \phi = .080$
nicht erwerbstätig	165 (28.6%)	31 (22.3%)	87 (30.3%)	18 (32.1%)	29 (30.9%)	

χ^2 : Chi²-Test; p : Signifikanz; ϕ : Effektstärke Phi. Die Signifikanzgrenzen wurden auf dem 10%- (°), dem 5%- (*), dem 1%- (**), und dem 0.1%- Niveau (***) festgelegt. Die Effektstärkenmaße werden als kleine (+), mittlere (++) und große (+++) Effektstärken interpretiert.

Anhang E: Patienteninformation, Einverständniserklärung, Katamnese-Anschreiben

1. Patienteninformation
2. Einverständniserklärung
3. Anschreiben zur Katamnese
4. Erinnerungsschreiben zur Katamnese



Institut und Poliklinik für Medizinische Psychologie
Martinistr. 52
20246 Hamburg

Kliniklogo und Klinikadresse

Ansprechpartnerinnen:
Dipl.-Psych. H. Melchior; Tel.: 040 - 42803 - 3871
Email: h.melchior@uke.uni-hamburg.de

*Ansprechpartner/in in der Klinik:
Name, Dienstbezeichnung, Telefonnummer*

Dipl. Soz. K. Hofreuter-Gätgens; Tel.: 040 – 42803 - 2001
Email: hofreuter@uke.uni-hamburg.de

Patienteninformation zur Studie:

"Optimierung der stationären Behandlung von Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen:

Bestimmung der individualisierten Mindestdosis psychotherapeutischer Maßnahmen und Analyse von Aspekten sozialer Ungleichheit“

Sehr geehrte Patientin, sehr geehrter Patient,

Sie befinden sich derzeit zu einer stationären Behandlung in der Curtius Klinik Bad Malente. Wir bemühen uns, Ihnen alle notwendigen und geeigneten Behandlungen zukommen zu lassen, die zur Besserung Ihres seelischen und körperlichen Befindens beitragen. Wir sind ständig bemüht, unsere Behandlung den neuesten Erkenntnissen anzupassen und auch einen eigenen Beitrag hierzu zu leisten.

Deshalb führen wir gemeinsam mit dem Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf (Prof. Dr. Dr. Uwe Koch) eine wissenschaftliche Studie zum Zusammenhang von Dauer und Wirksamkeit psychotherapeutischer Maßnahmen in der stationären Behandlung durch. Zielsetzung dieses Projekts ist es, zu untersuchen, wie sich Ihr Befinden und Ihr Gesundheitszustand im Verlauf der Behandlung verändern. An der Studie können alle Patientinnen und Patienten, die zurzeit in der Curtius Klinik behandelt werden, teilnehmen. Die Studie soll die Wissensgrundlage zur Versorgung der Patientinnen und Patienten verbessern und damit ermöglichen, die Versorgung noch stärker an Ihren Bedürfnissen auszurichten.

Wir möchten Sie herzlich um Ihre Mithilfe bei der Durchführung dieser wissenschaftlichen Studie bitten.

Wenn Sie sich für eine Mitarbeit entscheiden, bekommen Sie zu Beginn und am Ende der stationären Behandlung sowie ein halbes Jahr danach einen Fragebogen. Die ersten beiden Fragebögen erhalten Sie hier in der Klinik, den dritten Fragebogen bekommen Sie nach einem

halben Jahr per Post zugeschickt. Das Ausfüllen des Fragebogens dauert jeweils etwa 20-30 Minuten. Weiterhin erhalten Sie während Ihres Klinikaufenthalts wöchentlich jeweils einen Kurzfragebogen. Um Sie möglichst wenig zu belasten, ist dieser Fragebogen sehr kurz gehalten, das Ausfüllen dauert etwa 5 bis 10 Minuten. Die wiederholte Befragung zu mehreren Zeitpunkten ist notwendig, um Veränderungen in Ihrer Befindlichkeit im Zeitverlauf genau zu erfassen. Zusätzlich werden Daten aus Ihrem Entlassbericht zu Ihrer Erkrankung (Diagnose, Entlassungsstatus) sowie zu Art und Umfang der hier durchgeführten therapeutischen Maßnahmen erhoben. Dabei werden alle personenbezogenen Angaben in dem Bericht von Mitarbeitern der Klinik geschwärzt, so dass für die Forschergruppe keine Rückschlüsse auf einzelne Personen möglich sind. Für diese Tätigkeit werden die zuständigen Klinikmitarbeiter von ihrer Schweigepflicht entbunden. Diese Daten dienen dazu, eine Aussage darüber zu bekommen, wie wirksam die Maßnahmen, die Sie in der Klinik erhalten haben, für Sie gewesen sind.

Im Folgenden ist der **konkrete Ablauf der Studie** beschrieben:

1. Wenn Sie zur Mitarbeit bereit sind, erhalten Sie zu **Beginn Ihres Klinikaufenthalts** die **Einwilligung zur Studienteilnahme** und einen ersten **Fragebogen**.
2. Unterschreiben Sie bitte die **Einwilligung zur Studienteilnahme**, und geben Sie diese der/dem für Sie zuständigen Mitarbeiterin/Mitarbeiter.
3. Füllen Sie bitte den **Fragebogen** aus, stecken ihn anschließend in den beiliegenden Umschlag, verschließen den Umschlag und geben den verschlossenen Umschlag ebenfalls der/dem für Sie zuständigen Mitarbeiterin/Mitarbeiter.
4. Am Ende der ersten Woche Ihres Aufenthalts in der Klinik erhalten Sie den ersten **Kurzfragebogen**. Diesen geben Sie bitte, vollständig ausgefüllt und in dem beiliegenden Umschlag verschlossen bei der/dem zuständigen Mitarbeiterin/Mitarbeiter ab. Dieser Vorgang wird **jede Woche erneut mit dem gleichen Ablauf wiederholt**. Es ist besonders wichtig, dass Sie den Kurzfragebogen jede Woche vollständig ausfüllen und abgeben. Nur so können wir über die gesamte Dauer Ihres Aufenthalts Veränderungen in Ihrem Befinden erfassen.
5. Zum **Ende Ihres Klinikaufenthaltes** erhalten Sie einen weiteren **Fragebogen**. Bitte füllen Sie auch diesen aus, verschließen ihn im beiliegenden Umschlag und händigen diesen vor Ihrer Abreise ebenfalls der/dem zuständigen Mitarbeiterin/Mitarbeiter aus.

6. Die Einwilligung mit Ihren Personalien verbleibt zur Dokumentation für die Dauer des Forschungsprojektes in der Klinik.

7. Etwa **6 Monate nach Beendigung Ihrer stationären Behandlung** wird Ihnen der letzte **Fragebogen** von unserer Klinik aus zugesandt. Diesen füllen Sie bitte ebenfalls aus und senden ihn in dem beigefügten adressierten Freiumschlag an das Institut für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf zurück. Die Auswertung Ihrer Fragebögen erfolgt ausschließlich am dortigen Institut.

Ihre Teilnahme ist selbstverständlich freiwillig und Sie können jederzeit, auch wenn Sie schon Ihre Einwilligung gegeben haben, aus der Studie ausscheiden. Falls Sie nicht an der Studie teilnehmen möchten, entstehen Ihnen keinerlei Nachteile. Im Falle eines Widerrufs Ihrer Einwilligung werden die bis dahin erfassten Daten gelöscht.

Um die verschiedenen Fragebögen zuordnen zu können, sind diese mit einer einheitlichen fortlaufenden Code-Nummer versehen. Die Bearbeitung Ihrer Angaben erfolgt pseudonymisiert, das heißt ohne Nennung Ihres Namens, Ihres Geburtsdatums, Ihrer Anschrift oder anderer Angaben, die Rückschlüsse auf Ihre Person zulassen. Damit sind die Bestimmungen des Bundesdatenschutzgesetzes erfüllt, eine Weitergabe der Daten an staatliche oder private Einrichtungen ist ausgeschlossen. Eine Zuordnungsliste von Patientendaten und Code-Nummern sowie die unterschriebene Einwilligung werden in der Klinik verwahrt. Beide Unterlagen werden von wenigen Mitarbeitern der Klinik verwaltet und werden nicht von den Wissenschaftlern/innen eingesehen. Nach Abschluss der Studie zum Ende des Jahres 2010 werden sämtliche personenbezogenen Daten von den beteiligten Wissenschaftlern/innen des Instituts für Medizinische Psychologie gelöscht.

Wir möchten Sie noch darauf hinweisen, dass die Untersuchung nicht auf Veranlassung der Kostenträger erfolgt und ausschließlich Forschungszwecken dient.

Wenn Sie noch Fragen haben, stehen wir Ihnen jetzt und auch jederzeit später gerne zur Verfügung.

Im Namen der Klinik möchte ich mich im Voraus für Ihre Unterstützung und Mitarbeit bedanken!

(Chefarzt/Chefärztin)

EINWILLIGUNG DES PATIENTEN

Name: _____

Vorname: _____

Adresse: _____

Ich bin über Inhalt und Zweck der Studie „Optimierung der stationären Behandlung von Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen: Bestimmung der individualisierten Mindestdosis psychotherapeutischer Maßnahmen und Analyse von Aspekten sozialer Ungleichheit“, die vom Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf durchgeführt und ausgewertet wird, informiert worden. Zu diesem Zweck habe ich ein Informationsschreiben erhalten.

Mir wurde versichert, dass keine personenbezogenen Angaben (Name, Geburtsdatum, Adresse oder sonstige Angaben, die Rückschlüsse auf meine Person zulassen) an Dritte weitergegeben werden und dass die in der Studie erhobenen Daten gelöscht werden, sobald sie für die weitere wissenschaftliche Auswertung nicht mehr erforderlich sind.

Ich möchte die Studie durch meine Beteiligung unterstützen und willige ein, Fragebögen, die mir ausgehändigt oder zugeschickt werden, auszufüllen.

Ich bin damit einverstanden, dass Daten aus dem Entlassbericht zur Erkrankung, zum Entlassstatus und zu den erhaltenen therapeutischen Maßnahmen in pseudonymisierter Form den Forschungsunterlagen zugeführt werden. Zur Weitergabe der pseudonymisierten Angaben werden die beteiligten Mitarbeiter der Klinik von ihrer Schweigepflicht entbunden.

Unter den in dem Informationsschreiben genannten Voraussetzungen erkläre ich meine Einwilligung zur Teilnahme an der Studie.

Datum und Unterschrift Patient

Der Patient wurde von mir über die Studie aufgeklärt.

An

«Anrede»

«Vorname» «Name»

«Adresse1»

«Postleitzahl» «Ort»

Ihre Teilnahme an der Studie „Optimierung der stationären Behandlung von Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen“

Sehr geehrte Frau «Name»,

vor etwa einem halben Jahr hatten Sie sich anlässlich Ihrer Behandlung in unserer Einrichtung bereit erklärt, an unserer wissenschaftlichen Studie „Optimierung der stationären Behandlung von Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen“ teilzunehmen, die wir gemeinsam mit einigen anderen Fachkliniken und dem Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf durchführen.

Ziel dieser Studie ist es, die Behandlung für Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen noch mehr an den Bedürfnissen der einzelnen Betroffenen auszurichten. Darüber hinaus geht es auch darum, den längerfristigen Verlauf im Alltag nach Ende ihrer stationären Behandlung zu untersuchen. Deshalb möchten wir Sie nun – wie seinerzeit bereits angekündigt – nach Ablauf eines halben Jahres um Ihre Mithilfe für die abschließende Befragung bitten. Damit der Ihnen entstehende Zeitaufwand so gering wie möglich gehalten wird, haben wir einen kürzeren Fragebogen konzipiert.

Um die verschiedenen, von Ihnen ausgefüllten Fragebögen zuordnen zu können, sind diese mit einer einheitlichen Code-Nummer versehen. Dadurch ist gewährleistet, dass bei der Datenverarbeitung im Institut für Medizinische Psychologie die ärztliche Schweigepflicht gewahrt wird. Die Bearbeitung Ihres Fragebogens erfolgt dort anonym, das heißt ohne Nennung Ihres

Namens, Ihrer Anschrift oder anderer Angaben, die Rückschlüsse auf Ihre Person zulassen. Eine Weitergabe der Daten an staatliche oder private Einrichtungen ist ausgeschlossen. Nach Abschluss der Studie werden sämtliche Sie betreffenden Daten gelöscht.

Wir möchten Sie nun bitten, den Fragebogen auszufüllen und diesen im beigelegten, vorfrankierten Briefumschlag an das Institut für Medizinische Psychologie in Hamburg zurück zu schicken.

Wir wünschen uns natürlich, dass möglichst alle Patientinnen und Patienten, die auch während Ihrer Behandlung in unserer Klinik an der Befragung teilgenommen haben, diesen letzten Fragebogen ausfüllen. Ihre Erfahrungen und Einschätzungen nach Beendigung der Behandlung sind ein wesentlicher Bestandteil unserer Untersuchung. Bitte denken Sie daran, dass wir nur mit **Ihrer** Hilfe eine wissenschaftliche Aussage über die Behandlung treffen können.

Vielen Dank für Ihre Unterstützung!

Mit freundlichen Grüßen,

(Chefarzt/Chefärztin)

Bei Fragen wenden Sie sich bitte an:

Frau Dipl.-Psych. Hanne Melchior
Universitätsklinikum Hamburg-Eppendorf
Institut und Poliklinik für Medizinische Psychologie
Martinistr. 52
20246 Hamburg

Tel.: 040-7410-53871
Email: h.melchior@uke.uni-hamburg.de

An
«Anrede»
«**Vorname**» «**Name**»
«Adresse1»

«Postleitzahl» «Ort»

Sehr geehrte Frau «Name»,

vor einigen Wochen hatten wir Ihnen einen Fragebogen aus unserer Studie "Optimierung der stationären Behandlung von Patientinnen und Patienten mit psychischen/psychosomatischen Erkrankungen" zugesandt, die wir gemeinsam mit dem Institut für Medizinische Psychologie des Universitätsklinikums Hamburg-Eppendorf durchführen.

Wir möchten Sie auf diesem Wege noch einmal an diese Studie erinnern und Sie bitten, den Fragebogen auszufüllen, da dies sehr wichtig für die wissenschaftliche Auswertung ist. Nur wenn uns viele Patientinnen und Patienten antworten, besteht die Chance, zu einer realistischen Einschätzung der Wirksamkeit der stationären Behandlung zu gelangen. Die Teilnahme an unserer Studie ist freiwillig und wie zugesichert, dienen alle Ihre Angaben ausschließlich Forschungszwecken und werden nicht an Dritte weitergeleitet.

Wir würden uns sehr freuen, wenn Sie den ausgefüllten Fragebogen in dem zugehörigen Rückumschlag an das Institut für Medizinische Psychologie zurücksenden. **Denn gerade Ihre Meinung ist uns besonders wichtig.** Nur gemeinsam mit **Ihren Antworten** können wir uns bemühen die Behandlung zu verbessern.

Wenn Sie Fragen zur Studie oder zum Ausfüllen des Fragebogens haben, können Sie sich selbstverständlich gerne an das Institut für Medizinische Psychologie unter der Telefonnummer 040 / 7410 - 53871 (Frau Dipl.-Psych. H. Melchior) wenden.

Sollten Sie den Fragebogen zwischenzeitlich verschickt haben, bedanken wir uns sehr herzlich für Ihre Mitarbeit und bitten Sie, dieses Schreiben nicht weiter zu beachten.

Herzlichen Dank für Ihre Unterstützung unserer Studie!

Mit freundlichen Grüßen

(Chefarzt/Chefärztin)