

Wie häufig ist therapeutischer Misserfolg in der ambulanten Psychotherapie?

Ergebnisse aus einer verhaltenstherapeutischen Hochschulambulanz

Frank Jacobi^{1,2}, Stefan Uhmann¹ und Jürgen Hoyer¹

¹Technische Universität Dresden, Klinische Psychologie und Psychotherapie

²Psychologische Hochschule Berlin, Klinische Psychologie (Schwerpunkt Verhaltenstherapie)

Zusammenfassung. Für $N = 1776$ Psychotherapien einer verhaltenstherapeutischen Hochschulambulanz ermittelten wir die Raten für Nonresponse, Verschlechterung und unbefriedigende Therapieergebnisse. Für eine Teilstichprobe berichten wir Abbruchraten für verschiedene Therapiephasen. Outcome-Indikatoren waren die Therapieerfolgs-Einschätzung von Therapeut und Patient, das globale Funktionsniveau (Therapeuteinschätzung) und zwei störungsübergreifende Fragebogenmaße zur allgemeinen Symptombelastung und zur Depressivität. Die verschiedenen Maße weisen unterschiedliche (Miss-)Erfolgsraten auf, Verschlechterungen sind generell selten (0,8–4,3%). Ein kriterienbezogener Therapie-Misserfolg – eine (reliable) Verschlechterung oder ein hoch-symptomatischer Abschlusswert (eine Standardabweichung über dem durchschnittlichen Prä-Wert) – ist bei 11,2% der Stichprobe zu verzeichnen. Misserfolg korrelierte mit dem Schweregrad zu Therapiebeginn, nicht aber mit der Therapieerfahrung des Therapeuten. Wir diskutieren die Verschlechterungsraten und stellen heraus, dass die Misserfolgs- gegenüber der Erfolgsperspektive eine vertiefende Bewertung von therapeutischen Maßnahmen ermöglicht. Schlüsselwörter: therapeutischer Misserfolg, naturalistische Psychotherapieforschung, Phase-IV-Forschung, Evaluationsmethoden

How frequent is negative outcome in outpatient psychotherapy? Results from a CBT outpatient clinic

Abstract. We analysed the rate of non-response, deterioration, and unfavourable treatment outcomes for $N = 1,776$ psychotherapies in a cognitive-behavioural outpatient clinic. In a sub-sample, the frequency of drop-out during different treatment phases (treatment request, diagnostics, preparatory sessions, active therapy) was also examined. The subjective treatment evaluation of therapist and patient, the global assessment of functioning (as rated by therapists), and two global measures of symptom load and depression served as outcome indicators. Rates of negative outcome varied depending on the measure used. Deterioration was rare, regardless of measure (0,8–4,3%). Negative treatment outcome (defined as reliable deterioration in one of the measures or critically high symptom score at post-treatment) was seen in 11,2% of the sample. Negative outcome correlated positively with symptom severity at beginning of treatment, but not with the therapist's experience. We discuss the rates of deterioration and emphasize the incremental value of reflecting treatment failure in a comprehensive evaluation of therapeutic methods.

Key words: negative outcome in psychotherapy, naturalistic psychotherapy research, Phase-IV research, psychotherapy evaluation

Theoretischer Hintergrund

Die Wirksamkeit psychotherapeutischer Verfahren ist durch eine Vielzahl randomisiert-kontrollierter Studien (RCT) gut belegt (Lambert & Ogles, 2004). Diese Ergebnisse der Phase-III-Forschung legitimieren Psychotherapie in der Versorgung psychischer Störungen. Sie sollten durch naturalistische Anwendungsstudien im Versorgungsalltag (Phase-IV-Forschung) ergänzt werden (Übersicht: Jacobi, 2006). Bei Studien dieses Typs interessiert die Bewährung einer Therapie in der Alltagspraxis und die

Vergleichbarkeit der Ergebnisse mit denen aus RCTs (Hiller, Schindler, Andor & Rist, 2011). Der Fokus solcher naturalistischen Studien an großen unselegierten Stichproben liegt auf der klinischen Signifikanz der beobachteten Veränderungen (Ogles, Lunnen & Bonesteel, 2001). Beide Analyseformen, also jene zur Wirksamkeit (*efficacy*) genauso wie jene zur Effektivität (*effectiveness*) klinisch-psychologischer Interventionen, betrachten in der Regel aber lediglich, ob und wie stark sich Patienten hinsichtlich ihrer Symptome und Beeinträchtigungen gegenüber einer Vergleichsgruppe oder gegenüber ihren Ausgangswerten verbessern.

Für die Versorgungspraxis und die vergleichende Bewertung von Psychotherapieverfahren ist aber noch ein anderer Aspekt relevant: Wie hoch ist der Anteil derjenigen Patienten, bei denen keine Besserung erreicht wird oder es sogar zu Verschlechterungen kommt? Die Folgen unerwünschter Psychotherapieeffekte und Fehlentwicklungen in der Psychotherapie können für die Betroffenen gravierend sein. Vor dem Hintergrund des ethischen Grundprinzips der Schadensvermeidung (Noll-Hussong, 2011) und der Notwendigkeit der Legitimation des Einsatzes klinisch-psychologischer Interventionen (Barlow, 2004) ist es ein Zeichen von Professionalisierung, sich mit Risiken und Nebenwirkungen in der Psychotherapie explizit zu befassen. Dieser Bereich wurde in der systematischen Psychotherapieforschung schon früh thematisiert (Bergin, 1963), im Verhältnis zur Therapieerfolgs-Perspektive aber vernachlässigt (Übersichten z. B. bei Barlow, 2010; Bergin, 1963; Foa & Emmelkamp, 1983; Hoffmann, Rudolf & Strauss, 2008; Jacobi, 2001; Märten & Petzold, 2002; Mohr, 1995). Misserfolge und Verschlechterungen lassen sich nicht einfach auf den Kehrwert der Verbesserungen reduzieren. In seinem einflussreichen Überblicksartikel „*Psychological treatments that cause harm*“ stellt Lilienfeld (2007) heraus, dass es angesichts des gesicherten Wirksamkeitsnachweises vieler psychotherapeutischer Interventionen („*empirically supported treatments*“) ebenso wichtig sei, Verschlechterungen bis hin zu möglichen Therapieschäden zu betrachten. Letztere würden zwischen Therapieverfahren sogar besser differenzieren als Wirksamkeitsnachweise. Ferner bestehe in der Reduktion von Misserfolg und Verschlechterung das größte Potenzial für die Weiterentwicklung und Differenzierung psychotherapeutischer Maßnahmen.

Trotz der großen Zahl kontrollierter Therapiestudien sind Daten zu therapeutischem Misserfolg in der psychotherapeutischen Praxis rar. Therapieeffektmaße anhand von Mittelwertvergleichen, z. B. Effektstärken, haben nämlich den Nachteil, dass Informationen über individuelle Veränderungen verloren gehen: „*All too often negative responders are not identified at all – frequently they are simply buried in the outcome variance*“ (Mohr, 1995, S. 21). Geeigneter sind Ansätze, die für *jeden Patienten* angeben, ob und welche klinisch relevante Veränderung erfolgt ist.

Übersichtsarbeiten beziffern die Raten für Verschlechterungen (*deterioration*) in der Psychotherapie mit 5–15% und für ausbleibende Verbesserungen (*nonresponse*) mit 10–50% (z. B. Fischer-Klepsch, Münchau & Hand, 2009; Lambert, Hawkins & Hafield, 2002; Lambert & Ogles, 2004; Mohr, 1995). Diese Eckwerte beziehen sich aber in der Regel nur auf die Population derer, die eine Psychotherapie abschließen und auch am Ende der Therapie bereit sind, Rückmeldung zu geben. Um therapeutische Maßnahmen und ihren Erfolg umfassend bewerten zu können, ist es darüber hinaus wichtig, auch Misserfolge *im Vorfeld* der Therapie zu erfassen (wenn z. B. die Behandlung trotz Indikation nicht zustande kommt) und *Therapieabbrüche* (Lincoln, Rief, Hahlweg, Frank, von Witzleben, Schroeder et al., 2005; Cinkaya, Schindler & Hiller, 2011; Wierzbicki &

Pekarik, 1993) ebenso einzubeziehen wie „versteckte“ Misserfolge (bei fehlenden Nachuntersuchungen bzw. nicht dokumentierten Abschlusswerten, Hiller et al., 2011).

In der vorliegenden Studie soll die Häufigkeit von ausbleibendem Therapieerfolg und Verschlechterungen für das naturalistische Setting einer verhaltenstherapeutischen Ambulanz umfassend beschrieben werden. Im Einzelnen sollen folgende Fragen beantwortet werden:

1. Wie häufig sind Therapieabbrüche in verschiedenen Abschnitten des Therapieprozesses (Erstkontakt und Aufnahme auf Warteliste, Probatorik, eigentliche Therapie nach Kassenantrag)?

2. Wie häufig sind bei denjenigen, die in Therapie gelangen, Prä-Post-Misserfolge in verschiedenen Veränderungsdimensionen? Dabei unterscheiden wir zwischen Nonresponse, schlechtem Abschlusswert oder Verschlechterung jeweils hinsichtlich a) globaler Urteile, b) allgemeiner Symptombelastung, c) Depressivität und d) generellem Funktionsniveau. Ferner analysieren wir, ob fehlende Post-Werte systematisch mit Misserfolg assoziiert sind.

3. Welche Korrelate reliabler Verschlechterungen bzw. schlechter Abschlusswerte lassen sich identifizieren? Hierbei ist patientenspezifisch (Soziodemografie, Diagnosen, Schweregrad), behandlungsspezifisch (Therapieumfang) und therapeutenspezifisch (Therapieerfahrung) zu differenzieren.

Methoden

Setting und Stichprobe

Die Berechnungen werden auf der Basis von Dokumentationsdaten der Dresdner Institutsambulanz für Psychotherapie (IAP) des Zeitraums 2004–2010 vorgenommen. In dieser Hochschulambulanz wird Richtlinienpsychotherapie (Verhaltenstherapie) überwiegend von Ausbildungskandidaten in der Weiterbildung zum Psychologischen Psychotherapeuten, aber auch von approbierten Psychotherapeuten durchgeführt.

Die Diagnostik erfolgt mehrstufig (vgl. z. B. Hoyer, Uhlmann, Rambow & Jacobi, 2009): (1) diagnostisches Interview (Composite International Diagnostic Interview, CIDI, Wittchen & Pfister, 1997; vgl. auch Hoyer, Ruhl, Scholz & Wittchen, 2006) und Fragebogendiagnostik mit u. a. Brief Symptom Inventory (BSI; Franke, 2000) und Beck-Depressions-Inventar (BDI; Hautzinger, Bailer, Worrall & Keller, 1995); (2) Bewertung der Ergebnisse und Rückmeldung an Patienten durch klinisch erfahrene Psychotherapeuten; (3) im Rahmen der Probatorik durch behandelnde Therapeuten u. a. Vergabe von Behandlungsdiagnosen im Kassenantrag und Einschätzung der allgemeinen Funktionsfähigkeit des Patienten gemäß Achse V des DSM-IV (Global Assessment of Functioning, GAF, American Psychiatric Association, 2000). Für die

Tabelle 1. Kennwerte der Stichprobe und Therapien ($N = 1776$)

Alter ($M = 35.62, SD = 13.51$) (%)		Komorbidität: Anzahl Behandlungsdiagnosen ($M = 1.70, SD = 0.88$) (%)	
<18	1.5	1	51.4
18 – 34	53.2	2 – 3	44.5
35 – 49	28.3	4 – 5	4.1
50 – 64	13.5		
65+	3.5		
		Symptombelastung, Depressivität und allgemeines Funktionsniveau (Prä) (%)	
Geschlecht (%)		BSI ($M = 1.05, SD = 0.63$)	
weiblich	69.7	≤ 0.56	25.5
männlich	30.3	> 0.56	74.5
		> 1.68	16.0
Schulbildung (%)		BDI ($M = 16.91, SD = 9.75$)	
kein Schulabschluss	1.8	≤ 14.29	45.5
Hauptschule	9.9	> 14.29	54.5
Realschule	33.1	> 26.65	16.2
Abitur	55.1		
		GAF ($M = 63.35, SD = 12.54$)	
		≥ 70	39.7
		< 70	60.3
		< 51	16.7
Behandlungsdiagnosen (primär) (%)		Therapiecharakteristika	
Major Depression (F32/33)	26.0	Anzahl BE ($M = 31.52, SD = 15.89$) (%)	
Dysthymie	3.6	≤ 25	51.6
Panikstörung/Agoraphobie	16.4	26 – 45	20.4
Soziale Phobie	13.1	46 – 59	21.2
spezifische Phobien	3.3	60+	6.8
Generalisierte Angststörung	3.9		
Zwangsstörung	3.7	Erfahrung Therapeut (%)	
Posttraumatische Belastungsstörung	6.0	< 300 BE	64.4
F43.x außer PTSD	4.1	≥ 300 BE	18.9
somatoforme Störung	2.9	approbiert	16.8
Essstörung	9.2		
Persönlichkeitsstörung	1.3		
sonstige	6.5		

Anmerkungen: M : Mittelwert, SD : Standardabweichung, BSI: Brief Symptom Inventory; BDI: Beck-Depressions-Inventar; GAF: Global Assessment of Functioning; BE: Behandlungseinheiten.

wissenschaftliche Verwendung der Daten geben Patienten bereits vor dem diagnostischen Interview ihre informierte Einwilligung. Die Zuordnung von Fällen zu Therapeuten erfolgt quasi zufällig, da der Fall an erster Position der Warteliste dem Therapeuten zugewiesen wird, der einen Fall übernehmen will. Nur in Ausnahmefällen (<5%) wird diese Regel gebrochen, z. B. wenn eine ausreichende gute Prognose nur bei Einsatz eines erfahrenen Therapeuten getroffen werden kann.

Zu Therapieende werden im Rahmen der Abschlussuntersuchung unter anderem von Therapeut und Patient globale Urteile zum Therapieerfolg abgegeben (CGI; 7-stufige Likert-Skala mit 1/sehr deutlich gebessert, 2/deutlich gebessert, 3/etwas gebessert, 4/keine Veränderung, 5/etwas verschlechtert, 6/deutlich verschlechtert, 7/sehr deutlich verschlechtert), sowie die Eingangsfragebögen erneut vorgegeben.

Einschlusskriterium für die vorliegenden Analysen ($N = 1776$) waren: (1) Vorliegen des globalen Therapeutenurteils bei Therapieende und (2) regulärer Beginn der Therapien (d. h. vollständig durchlaufene Probatorik-Phase und erfolgreicher Kassenantrag).

Lag für die Berechnung der Erfolgskategorien von Patientenseite kein Post-Wert vor (Patient füllte Abschlussfragebögen nicht oder nur unvollständig aus; Häufigkeit: globales Urteil Patient: $N = 482$ bzw. 27.1%, BSI: $N = 546$ bzw. 30.7%, BDI: $N = 563$ bzw. 31.7%), aber ein 6-Monats-Katamnesewert, so wurde letzterer anstelle des fehlenden Post-Wertes verwendet. Bei diesem Vorgehen lagen die Rücklaufquoten bei 72.2% (BSI, 53 fehlende Post-Werte durch Katamnese-Werte ersetzt) bzw. 71.4% (BDI, 55 fehlende Werte ersetzt). Lagen weder Post- noch Katamnesewert, aber eine Verlaufsmessung vor (zur 25. Behandlungseinheit, nur bei Langzeittherapien), so wurde diese (im Sinne der *last observation carried forward*) anstelle des fehlenden Post-Wertes verwendet. Dies erhöhte die Rücklaufquoten auf 76.0% (BSI, 66 fehlende Post-Werte durch Verlaufs-Werte ersetzt) bzw. 75.3% (BDI, 70 fehlende Werte ersetzt). Diese Quoten liegen niedriger als der methodisch erwünschte Idealwert von über 90% (Hiller, Bleichhardt & Schindler, 2009). Daher soll eine gesonderte Missing-Analyse vorgenommen werden, um zu bestimmen, inwieweit bei den Patienten mit fehlenden Post-Werten eine relativ höhere Quote an schlechten Abschlusswerten und reliabler Verschlechterung gegeben ist.

In Tabelle 1 sind die Stichprobenkennwerte aufgeführt. 69.7% der Patienten waren weiblich; das Durchschnittsalter betrug 35.62 Jahre ($SD = 13.51$), schwerpunktmäßig im Altersbereich von 18–34 (53.2%). Es findet sich ein für urbane Regionen und diese Altersverteilung typischer, relativ hoher Anteil von Patienten mit Abitur (55.1%). Diagnosen wurden vornehmlich im Bereich der depressiven und der Angststörungen vergeben. Bei der in Tabelle 1 aufgeführten Verteilung der durch die behandelnden Therapeuten erstellten Hauptdiagnosen ist zu beachten, dass Patienten meist die Kriterien für mehrere Behandlungsdiagnosen (durchschnittliche Diagnosezahl: 1.70) erfüllten. Ferner ist der Anteil an Persönlichkeitsstörungen aufgrund der Spezialisierungen der Einrichtung auf Achse I-Störungen möglicherweise geringer als sonst in der Versorgung.

Der Mittelwert im BSI lag bei $M = 1.05$ (Standardabweichung: $SD = 0.63$), im BDI bei $M = 16.91$ ($SD = 9.75$). Aufgrund der mangelnden Spezifität dieser Maße (Hoyer, Becker, Neumer, Soeder & Margraf, 2002) lag nur ein Teil der Stichprobe über deren klinischen Cut-Off-Werten (Werte bei Hiller et al., 2011): im BSI etwa 75%, im BDI etwa 55% und im GAF 60%. In einem hoch symptomatischen Bereich (definiert als mindestens eine SD schlechter als der durchschnittliche Prä-Wert) lagen – entsprechend der Normalverteilungsannahme – jeweils etwa 16% der Stichprobe.

Das Verhältnis von Kurzzeit- (bis zu 25 Sitzungen) zu Langzeittherapien (26 Sitzungen und mehr) beträgt 51.6% zu 48.4% bei einem mittleren Therapieumfang von 31.47 Behandlungseinheiten ($SD = 15.91$).

Therapeutische Behandlungserfahrung wurde über den Fortschritt der praktischen Ausbildung des Therapeuten zu Therapiebeginn operationalisiert und damit jedem Fall eine von drei Kategorien zugeordnet: (1) geringe Erfahrung: unter 300 ambulante Behandlungseinheiten (BE, 64.3%) im Rahmen der Psychotherapieausbildung (ohne Berücksichtigung stationärer oder anderweitiger psychotherapeutischer Tätigkeit); (2) moderate Erfahrung: mindestens 300 BE (18.8%); (3) professionelle Erfahrung: Approbation liegt vor (16.8%).

Nicht-Aufnahme von Therapie sowie frühzeitige und spätere Therapieabbrüche

Bereits im Vorfeld (Anmeldung, Probatorik) findet eine deutliche Auslese der Patienten statt. Diesen Aspekt quantifizieren wir, ebenso wie den der Therapieabbrüche, auf der Basis einer Teilstichprobe, welche in einer früheren Arbeit speziell hinsichtlich der Therapieabbrüche analysiert worden war (Hansche, 2008; s. u., Abb. 1), da die mit den Kodierungen der Abbrüche verbundenen aufwändigen Aktenanalysen nicht bei allen Patienten vorgenommen wurden. Diese Teilstichprobe war in allen o. a. Parametern mit der Reststichprobe vergleichbar. Bei den Abbrüchen im Verlauf einer regulären Therapie (also nach probatorischer Phase) wurde bei den dokumentierten Abbrüchen unterschieden in:

(1) unproblematische, z. B. beendet Patient die Therapie bei deutlicher Besserung der Symptomatik bzw. aufgrund neutraler Umstände, wie z. B. Wohnortwechsel oder von Störung unabhängiger Erkrankung, oder Patient wird weiterverwiesen, und

(2) problematische, z. B. organisatorische Probleme, z. B. keine Kostenübernahme oder terminliche Probleme patientenseitig, behandlungsbedingte Schwierigkeiten, wie z. B. problematische Therapeut-Klient-Beziehung, Abbruch seitens Therapeut, z. B. aufgrund als unzureichend eingeschätzter Therapiemotivation, oder keine Erreichbarkeit des Patienten mehr nach mehreren unentschuldigten Fehlterminen.

Diese Kategorisierung unterscheidet sich in Details von der später von Hiller und Kollegen (2011) veröffentlichten, nimmt aber im selben Sinne eine Unterscheidung in nicht- und qualitätsrelevante Abbrüche vor; problematisch bzw. qualitätsrelevant sind diejenigen Abbrüche, die bei der Beurteilungen der Qualität der Behandlungen und bei Überlegungen zu Verbesserungen berücksichtigt werden sollten.

Definition der (Miss-)Erfolgskategorien

Die globalen Urteile von Therapeut und Patient werden in den Kategorien „sehr stark gebessert“, „stark/etwas gebessert“, „unverändert“ und „verschlechtert“ zusammengefasst berichtet. Die Bildung der (Miss-)Erfolgskategorien der anderen Maße ist Tabelle 2 zu entnehmen. Die Intervalle für reliable Veränderungen (Reliable Change Index, RCI, Jacobson & Truax, 1991, BSI: ± 0.63 , BDI: ± 8.46) sowie die Cut-Offs zwischen dysfunktionalen vs. funktionalen Bereichen (BSI: 0.56, BDI: 14.29) wurden gemäß der Vorschläge von Hiller und Kollegen (2011) festgelegt. Für die Einschätzung des Therapeuten hinsichtlich der allgemeinen Funktionsfähigkeit (GAF) liegen keine entsprechenden Normen vor; wir verwenden hier als klinischen Bereich Werte von kleiner als 70 und als kritische Differenz für reliable Veränderung ein Intervall von 12.5 Punkten (entspricht einer SD in den GAF-Prä-Werten der Gesamtstichprobe).

Es wurde unterschieden zwischen a) Remission (*remission, recovery*): reliable Veränderung vom klinisch auffälligen in den funktionalen Bereich, b) Response: reliable Verbesserung, aber Abschlusswert noch nicht im funktionalen Bereich, c) Nonresponse: keine reliable Veränderung, sowie d) reliable Verschlechterung (*deterioration*). Fälle mit Prä-Werten im unauffälligen Bereich wurden nur bei den Berechnungen zu Verschlechterungen einbezogen. Solche Fälle mit nicht-klinischen Prä-Werten repräsentieren z. B. Patienten mit spezifischen psychischen Störungen (z. B. Phobien), deren Symptome in den Instrumenten BSI, BDI und GAF nicht zu einem Überschreiten des Cut-Offs führen.

Jenseits (reliabler) Veränderung sind Fälle mit klinisch auffälligen Post-Werten erklärungsbedürftig. Zusätzlich berichten wir daher die Rate der Fälle mit Abschlusswert über dem Cut-Off, sowie die Fälle mit einem hoch sympto-

Tabelle 2. Ergebniskategorien

	Remission: reliable Veränderung vom klinischen in Normalbereich	Response: reliable positive Veränderung; Post immer noch im dysfunktionalen Bereich	Nonresponse: keine reliable Veränderung	Verschlechterung: reliable negative Veränderung
BSI (Prä-Post; Patient)	1. positive Veränderung ≥ 0.63 2. Prä ≥ 0.56 3. Post ≤ 0.56	1. positive Veränderung ≥ 0.63 2. Prä und Post ≥ 0.56	1. Prä ≥ 0.56 2. Veränderung ≤ 0.63	jegliche negative Veränderung ≥ 0.63
BDI (Prä-Post; Patient)	1. positive Veränderung ≥ 8.46 2. Prä ≥ 14.29 3. Post ≤ 14.29	1. positive Veränderung ≥ 8.46 2. Prä und Post ≥ 14.29	1. Prä ≥ 14.29 2. Veränderung ≤ 8.46	jegliche negative Veränderung ≥ 8.46
GAF (Prä-Post; Therapeut)	1. positive Veränderung ≥ 12.5 2. Prä ≤ 70 3. Post ≥ 70	1. positive Veränderung ≥ 12.5 2. Prä und Post ≤ 70	1. Prä ≤ 70 2. Veränderung ≤ 12.5	jegliche negative Veränderung ≥ 12.5

Anmerkungen: BSI: Brief Symptom Inventory; BDI: Beck-Depressions-Inventar; GAF: Global Assessment of Functioning.

matischen Abschlusswert (definiert als mindestens eine SD schlechter als der durchschnittliche Prä-Wert der Stichprobe; BSI > 1.68, BDI > 26.65, GAF < 51), der akuten Behandlungsbedarf nahelegt.

Da die verschiedenen Methoden zur Operationalisierung von (Miss-) Erfolg verschiedene Häufigkeiten liefern, erscheint eine Kombination bei der Untersuchung von Korrelaten sinnvoll. Eine solche aggregierte Misserfolgs-Variable wurde daher operationalisiert über zwei hinreichende Kriterien: a) eine reliable Verschlechterung oder b) ein hoch symptomatischer Abschlusswert – jeweils in mindestens einem der einbezogenen Maße.

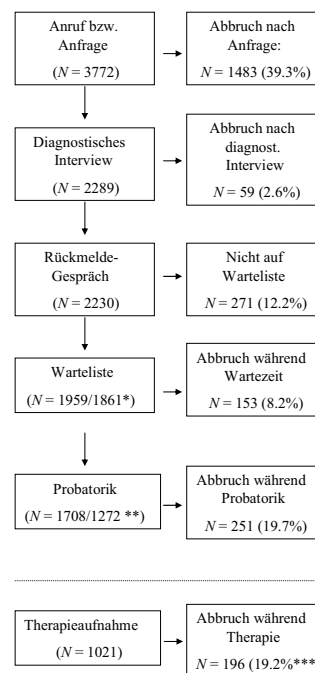
Statistische Auswertung

Wir verglichen die Raten verschiedener Abbruchs- und Erfolgskategorien miteinander und testeten diese mittels χ^2 -Test auf signifikante Zusammenhänge mit soziodemografischen und Therapievariablen. Unterschiede zwischen Teilstichproben bezüglich metrischer Variablen wurden mit Welch's t-Tests (für heterogene Varianzen) auf Signifikanz geprüft; bei ordinalem Datenniveau (Anzahl komorbider Diagnosen) wurde der Wilcoxon-Rangsummentest eingesetzt.

Ergebnisse

Therapieaufnahme und Therapieabbruch

Bei fast 40% derjenigen, die eine Therapieanfrage an die Behandlungseinrichtung richteten, blieb es bei der Anfra-



Anmerkungen: * N = 98 von dieser Analyse ausgeschlossen, da immer noch auf Warteliste, ** N = 436 von dieser Analyse ausgeschlossen, da Therapie noch nicht abgeschlossen, *** unter diesen N = 196 befanden sich N = 75 (38%) als „unproblematisch“ klassifizierte Abbrüche; somit beträgt die Quote problematischer Therapieabbrüche in diesen N = 1021 Therapien 11.9%.

Abbildung 1. Abbrüche im Vorfeld der eigentlichen Behandlungsaufnahme und nach Therapiebeginn (Teilstichprobe; Hansche, 2008); Prozentangaben zeilenbezogen.

ge (vgl. Abbildung 1). Knapp 3% brachen nach dem diagnostischen Eingangsinterview ab, und 12% wurden nach dem Rückmeldegespräch nicht auf die Warteliste genommen (23% ohne hinreichende Indikation, 54% weiterverwiesen, 7% ohne Therapiewunsch und 16% wegen anderer Gründe wie z. B. organisatorischer Probleme, Umzug etc.). Nach weiteren Abbrüchen während Wartezeit (8%) und Probatorik (20%) begannen insgesamt lediglich 32% derjenigen, die ursprünglich eine Anfrage gestellt hatten, nach erfolgreichem Kassenantrag die eigentliche Behandlung.

Therapieabbrüche nach Aufnahme der eigentlichen Behandlung (19.2%, $N = 196$ aus $N = 1021$ Therapien) verteilten sich wie folgt: a) Gegenüber 21.2% Therapieabbrüchen in der Kurzzeittherapie finden sich in den längeren Therapieformaten weniger Abbrüche (Langzeittherapie: 17.8%, Langzeittherapie mit mehr als 45 Behandlungseinheiten: 12.2%); b) von den 196 Abbrüchen nach begonnener Therapie wurden 75 als „unproblematisch“ klassifiziert, somit verblieben insgesamt 11.9% als „problematische“ Abbrüche ($N = 121$ aus $N = 1021$ Therapien).

Verteilung von (Miss-)Erfolg

Zur Identifizierung des Verhältnisses von Remission, Response, Nonresponse und Verschlechterung siehe zunächst Abbildung 2, in der die Verteilung der Prä- und Post-Werte am Beispiel des BDI dargestellt ist (vgl. Hiller et al., 2011; Jacobi, 2006). Der Cut-Off von 14.30 wird durch horizontale und vertikale Linien repräsentiert. Der Normalbereich befindet sich also links der entsprechenden Linie auf der Prä-Achse (Abszisse) bzw. unterhalb der entsprechenden Linie auf der Post-Achse (Ordinate).

Fälle mit Prä- und Post-Werten unterhalb des Cut-Offs wurden nicht abgebildet oder in Berechnungen einbezogen, da sie sich bereits Prä im Normalbereich befanden und zu Post nicht verschlechtert hatten (44.4% der Gesamtstichprobe, d.h. hier würden sich sonst fast ebenso viele Punkte befinden wie im Rest der Grafik). Fälle mit Nonresponse im Sinne keiner reliablen Veränderung machen einen Anteil von 27.1% aus; Remission 62.7%, Response 7.1% und reliable Verschlechterungen 3.0%. Reliable Verschlechterungen sind also relativ selten und Remissionen am häufigsten. Rund ein Drittel ist reliabel verbessert, aber Post noch immer im auffälligen Bereich, oder weist keine reliable Veränderung auf. Außerdem liegen zu Post noch 17.9% der Fälle im dysfunktionalen Bereich. Diese Verteilung entspricht einer Effektstärke (Mittelwertsdifferenz Prä-Post relativiert an Prä-SD) von $d = 0.97$.

Bei den globalen Urteilen (Tabelle 3) berichtet der Großteil sowohl der Therapeuten (88.9%) als auch der Patienten (92.7%) entweder „sehr starke“ oder „starke/etwas“ Verbesserung; keine Veränderung berichten 8.8% der Therapeuten (Patienten: 4.5%), und eine Verschlechterung 2.3% der Therapeuten (Patienten: 2.7%). Da bei $N = 482$ Patienten keine globale Einschätzung zu Therapieende vorlag und die Annahme besteht, dass fehlende Post-Werte mit Mißerfolg assoziiert sind, wurde eine zusätzliche Abschätzung auf Grundlage der Therapeutenurteile vorgenommen (s. u.).

Reliable Verschlechterungen sind auch in BSI, BDI und GAF (Tabelle 4) relativ selten (0.8–3.0%). Bezüglich der Fragebogenwerte betragen Response oder Remission zusammen knapp 50% (BSI) bis 70% (BDI); im GAF 67%.

Der Anteil derjenigen, die sich nach Therapie (noch) im klinisch auffälligen Bereich (entsprechend der o. g. Cut-Offs) befinden liegt bei 34.1% im BSI, 17.9% im BDI und 17.7% hinsichtlich des GAF. Hoch symptomatische Ab-

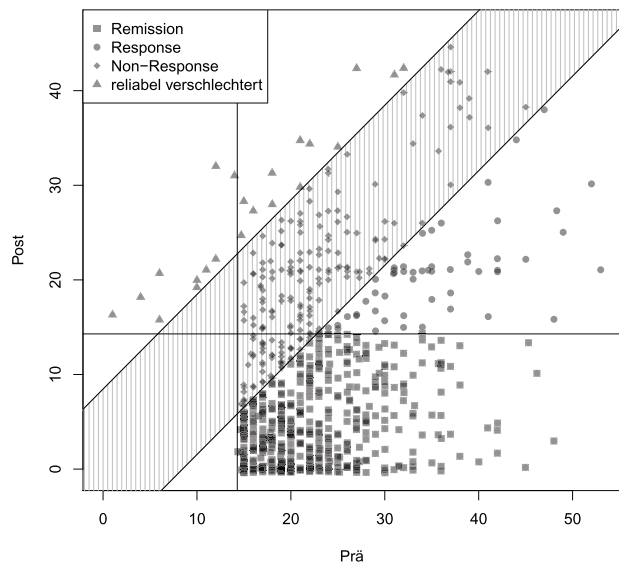


Abbildung 2. Prä-Post-Veränderungen im Beck-Depressions-Inventar (BDI) bei $N = 1328$ Patienten in ambulanter Verhaltenstherapie (Erläuterung: reliable Veränderungen befinden sich außerhalb der schraffierten Diagonale; diejenigen, die vor der Therapie einen unauffälligen Wert (< 14.29 ; $N = 600$) hatten, wurden nur bei reliabler Verschlechterung berücksichtigt); diese Verteilung entspricht einer Effektstärke von $d = 0.97$.

Tabelle 3. Prozentuale Verteilung der globalen Urteile Prä-Post von Therapeut und Patient

	„sehr stark gebessert“	„stark/etwas gebessert“	„unverändert“	„etwas/stark/sehr stark verschlechtert“
Therapeut	20.8	68.1	8.8	2.3
Patient	30.9 (26.4)	61.8 (60.1)	4.5 (8.7)	2.7 (4.1)

Anmerkungen: Urteil Therapeut: $N = 1776$, Urteil Patient: $N = 1294$; in Klammern: der ermittelte %-Wert, wenn nicht-vorliegende Post-Urteile auf Patientenseite ($N = 482$) unter Zuhilfenahme der vorliegenden Therapeutenurteile geschätzt werden.

Tabelle 4. Verteilung der Therapieergebnisse

	Remission [%]	Response [%]	Nonresponse [%]	Verschlechterung [%]	auffälliger Post-Wert [%]	hoch- symptomatischer Post-Wert [%]
BSI (Prä-Post)	37.9	10.7	48.6	2.8	34.1	5.7
BDI (Prä-Post)	62.7	7.1	27.1	3.0	17.9	3.6
GAF (Prä-Post)	61.3	5.2	32.8	0.8	17.7	4.5

Anmerkungen: BSI: Brief Symptom Inventory; BDI: Beck-Depressions-Inventar; GAF: Global Assessment of Functioning; Unterschiedliches N aufgrund von Missings (BSI: $N = 1339$; BDI: $N = 1328$; GAF: $N = 1748$); „auffälliger Post-Wert“: Post im dysfunktionalen Bereich (BSI > 0.56, BDI > 14.29, GAF < 70); „hoch-symptomatischer Post-Wert“: im Bereich von mehr als einer Standardabweichung schlechter als Durchschnitt Prä (BSI > 1.68, BDI > 26.65, GAF < 51); diejenigen, die einen Prä-Wert im funktionalen Bereich hatten (BSI: $N = 330$ bzw. 24.7%, BDI: $N = 600$ bzw. 45.1%, GAF: $N = 347$ bzw. 19.9%), wurden nur bei den Berechnungen zu Verschlechterungen einbezogen.

schlusswerte (1.0 SD über durchschnittlichem Prä-Wert) weisen beim BSI 5.7%, beim BDI 3.6% und beim GAF 4.5% der Patienten auf.

Sind fehlende Post-Werte systematisch mit Misserfolg assoziiert?

Patienten mit fehlenden Post-Werten haben im Vergleich zu denjenigen mit vorhandenen Post-Werten eine niedrigere Schulbildung (Hauptschulabschluss: 14% vs. 9%, Abitur: 49% vs. 57%; $\chi^2(1545) = 18.7, p < .001$) sowie höhere Schweregrade zum Prä-Zeitpunkt (Mittelwert BSI Prä: 1.13 vs. 1.02, $t(494.906) = 2.45, p < .05$; BDI Prä: 18.3 vs. 16.5, $t(498.463) = 2.79, p < .01$; GAF Prä: 61.4 vs. 63.9, $t(647.978) = -3.44, p < .001$). Da diese beiden Faktoren mit Misserfolg korrelieren (s. u.) legt dies nahe, dass Misserfolge bei denjenigen mit fehlendem Post-Wert häufiger sind als bei denjenigen mit vorhandenem Post-Wert. Ferner beurteilen die Therapeuten die Therapien, bei denen keine Post-Messung der Patienten vorliegt, deutlich schlechter ($\chi^2(3) = 128.8, p < .001$): 11% „sehr stark gebessert“ (vs. 24% bei denen mit Post-Messung), 64% „stark/etwas gebessert“ (vs. 69%), 21% „unverändert“ (vs. 5%), 4.3% Verschlechterung (vs. 1.6%). Das legt nahe, dass auch die entsprechenden Patientenurteile in dieser Subgruppe schlechter als in der Gesamtstichprobe ausfallen würden, wenn sie vorlägen.

Um hierfür zumindest für das Globalurteil der Patienten eine quantitative Abschätzung vorzunehmen, wurde angenommen, dass die Ergebniskategorien bei den fehlenden Patientenurteilen ebenso ungünstiger verteilt wären wie bei den Therapeutenurteilen (also nur etwa 0.5-mal so viel „sehr starke Besserung“ wie bei den vorliegenden Werten, 0.9-mal so viel „stark/etwas gebessert“, 4-mal so viel „unverändert“ und 3-mal so viel Verschlechterungen). Auf diese Weise gelangt man zu den Werten in Klammern bei den Patientenurteilen in Tabelle 3, d. h. die Abschätzung der wahren Werte für alle Patienten ergibt 26.4% Remission (vs. 30.9% beobachtet), 60.1% Response (vs. 61.8%), 8.7% Nonresponse (vs. 4.5%) und 4.1% Verschlechterungen (vs. 2.7%). Fehlende Fragebogenwerte können auf diese Weise nicht geschätzt werden, aber es

ist davon auszugehen, dass auch deren Ergebnisverteilung etwas ungünstiger liegen würde, wenn für alle Missings der wahre Wert bekannt wäre.

Korrelate von Misserfolg

Ein Therapie-Misserfolg im Sinne einer aggregierten Variable von Verschlechterung und/oder hoch symptomatischem Abschlusswert war bei 11.2% der Stichprobe zu verzeichnen. Keinerlei Zusammenhänge fanden sich zu Alter und Geschlecht des Patienten. Ein höherer Schulabschluss war mit einer geringeren Rate von Misserfolg assoziiert (Abitur: 7.9% vs. Realschule 15.0% und Hauptschule 17.0%; $\chi^2(1545) = 24.8, p < .001$).

Hinsichtlich klinischer Charakteristika zu Therapiebeginn war Misserfolg assoziiert mit Komorbidität ($W = 118934.5, p < .001$; bzw. bei einer Diagnose 7.4%, bei 2–3 Diagnosen 14.4% und bei 4–5 Diagnosen 25.0%; $\chi^2(1766) = 35.06, p < 0.001$) und schlechteren Prä-Werten (BSI: 1.44 vs. 0.99, $t(207.20) = -7.73, p < .001$; BDI: 23.1 vs. 16.1, $t(214.0) = -8.08, p < .001$; GAF: 54.5 vs. 64.5, $t(233.71) = 9.49, p < .001$). Bezüglich der Hauptdiagnosen gab es häufigere Misserfolge bei Posttraumatischer Belastungsstörung (20.8% vs. 10.6%, $\chi^2(1766) = 9.32, p < 0.01$). Weniger Misserfolge dagegen gab es unter der Hauptdiagnose Panikstörung/Agoraphobie (7.6% vs. 11.9%, $\chi^2(1766) = 4.16, p < 0.05$).

Keine signifikanten Unterschiede bei den Misserfolgsraten zeigten sich hinsichtlich der Therapieerfahrung der Therapeutinnen und Therapeuten (<300 BE: 11.8%, ≥ 300 BE: 10.2%, approbiert: 9.8%; $\chi^2(1760) = 1.31, p = 0.52$), wobei es zum Prä-Zeitpunkt keine Unterschiede im BDI, BSI oder GAF gab. Ferner dauerten Therapien, bei denen ein Misserfolg zu verzeichnen war, tendenziell länger (33.8 vs. 31.2 Behandlungseinheiten, $t(236.40) = -1.93, p = .054$).

Diskussion

Die vorliegende Arbeit berichtet den Anteil an Nonresponse, Verschlechterung und unbefriedigenden The-

rapieergebnissen bei der Abschlussmessung im Setting einer verhaltenstherapeutischen Ambulanz anhand einer konsekutiven Stichprobe von 1776 Therapien. Vor dem Hintergrund vorliegender repräsentativer naturalistischer Studien handelt es sich hinsichtlich Komorbidität und allgemeiner Belastung um eine klinisch relevant beeinträchtigte Patientenpopulation, wie sie üblicherweise in der ambulanten psychotherapeutischen Routineversorgung anzutreffen ist (vgl. Steffanowski, Kramer, Fembacher, Glahn, Bruckmayer, von Heymann et al., 2011; Wittmann, Lutz, Steffanowski, Kriz, Glahn, Völkle et al., 2011).

Zunächst wurde anhand einer Teilstichprobe gezeigt, dass nur etwa jede dritte Anfrage zur regulären Aufnahme einer Richtlinien-Psychotherapie führt. Bei den Patienten, bei denen eine reguläre Therapie zustande kam, lagen die Raten an Verschlechterungen in den verwendeten Maßen bei 0.8%–4.1% und an hoch symptomatischen Abschlusswerten bei 3.6%–5.7%. In Abhängigkeit vom Veränderungsmaß befinden sich im Vergleich zu den Verschlechterungen allerdings 18%–34% der Patienten am Ende der Therapie noch im dysfunktionalen Bereich (Therapiebeginn: 54%–80%). Therapie-Misserfolg definiert als irgendeine (reliable) Verschlechterung und/oder hoch symptomatischer Abschlusswert (mindestens eine *SD* schlechter als der Mittelwert der Gesamtgruppe zu Therapiebeginn in einem der eingesetzten Maße) war bei 11.2% der Therapien zu verzeichnen.

Nicht-Aufnahme bzw. Therapieabbruch

Hinsichtlich der die Gruppe derjenigen, die die Therapie nicht in Anspruch nehmen, bedarf es eigener Studien. Sie sind wichtig, um zu klären, ob bei dieser Gruppe eine Therapie indiziert bzw. aussichtsreich wäre, ob Barrieren für die Aufnahme einer Psychotherapie bestehen (z. B. aufgrund Stigmatisierung, vgl. Corrigan, 2004) und ob Veränderungen des Versorgungssystems notwendig wären, um einen sachgerechten Zugang zu den adäquaten Angeboten sicher zu stellen. In der vorliegenden Studie konnten jedoch lediglich diejenigen Fälle untersucht werden, in denen eine Therapie aufgenommen wird. Die innerhalb dieser Gruppe erfassten Therapieabbrüche (19%) sind mindestens mit einem globalen Therapeutenurteil sowie der Einschätzung des generellen Funktionsniveaus dokumentiert. Knapp 40% dieser Therapieabbrüche wurden allerdings als unproblematisch klassifiziert.

Wirkt die Verhaltenstherapie selten schädlich?

Für den Bereich der Psychotherapie existieren keine ausreichenden Vergleichsdaten oder Konventionen hinsichtlich der Frage, ob die hier berichteten Misserfolgsraten wirklich als „niedrig“ zu bezeichnen sind. Immerhin gilt aber, dass in mehreren Übersichtsarbeiten höhere Verschlechterungsraten von 5–15% berichtet wurden (Lambert & Ogles, 2004; Mohr, 1995).

Dass nur höchstens jede 25. Therapie von Therapeut und/oder Patient global als zumindest leichte Verschlech-

terung eingeschätzt wird, kann zum Teil auch mit selbstwertdienlichen, dissonanzreduzierenden oder auf einer Tendenz zu sozialer Erwünschtheit basierenden Verzerrungen interpretiert werden (Höfler, Gloster & Hoyer, 2010; Hoyer & Uhmann, 2008). Gleichwohl sind solche rückwirkenden Einschätzungen, ebenso wie Zufriedenheitsratings, psychologisch nicht ohne Bedeutung, denn sie stellen persönlich relevante sozial-kognitive Konstruktionen im Hinblick auf das subjektive Wohlbefinden dar. Im Rahmen einer Therapie-Evaluation haben diese Einschätzungen nach wie vor zu Recht eine prominente Stellung, auch wenn sie mit symptomatischen oder objektiven Veränderungen meist nur mäßig korrelieren (Lunnen, Ogles & Pappas, 2008). Die hier eingesetzten prospektiv angelegten Fragebogenmaße zeigen im Übrigen ähnlich niedrige Verschlechterungsraten, auch wenn sie für die Verzerrungen der retrospektiven Maße weniger anfällig sind.

Kann man aus dem Gesamtbild unserer Routinedaten den Schluss ziehen, dass die Verhaltenstherapie vergleichsweise selten schädlich wirkt? Es lassen sich einige Punkte anführen, die – zumindest für das Setting, aus dem die hier vorgestellten Daten stammen – dafür sprechen. Es handelt sich um auf relativ klar umschriebene Problemkonstellationen ausgerichtete Behandlungen, die zeitlich begrenzt sind. Ein Machtgefälle zwischen Therapeut und Patient ist immer gegeben, aber in der Verhaltenstherapie möglicherweise weniger stark ausgeprägt als bei anderen Therapieverfahren (Fäh, 2002; Tress & Erny, 2008; Wirth, 2005). In der Verhaltenstherapie ist die Bereitschaft bzw. Verpflichtung zur Einholung von Rückmeldung (therapiebegleitende Evaluation, Nutzung von Videodokumentation, Supervision/Intervision) hoch ausgeprägt. Ferner zählen der Einsatz von empirisch fundierten Vorgehensweisen und der Ethos, sich an wissenschaftlich geprüften Leitlinien zu orientieren, zu den Grundprinzipien der Verhaltenstherapie (Margraf & Lieb, 1995). Hierbei ist allerdings zu beachten, dass dies auch eine Quelle von schädlichem Therapeutenverhalten darstellen kann, das potenziell Misserfolge produziert: Wenn Therapeuten zu rigide an Therapiemanualen und vermeintlich allgemein gültigen empirisch gestützten Vorgehensweisen festhalten und damit darin unflexibel werden, ihre Maßnahmen und die therapeutische Beziehungsgestaltung auf komplexe Einzelfälle anzupassen, kann dies systematisch zu vermeidbaren Therapieabbrüchen und Fehlbehandlungen führen (Castonguay, Boswell, Constantino, Goldfried & Hill, 2010). Interessant ist aber in diesem Zusammenhang, dass unsere Ergebnisse darauf hinweisen, dass weniger erfahrene Therapeuten (bei denen man aufgrund von Unsicherheiten eher die Gefahr zu rigider Manualtreue erwarten könnte) sich in der Misserfolgsrate nicht von erfahrenen Therapeuten unterscheiden.

Methodische Einschränkungen

Es lagen zwar grundsätzlich für alle Therapien Therapeutenurteile vor, aber gerade unter der Misserfolgsperspektive ist eine niedrigere Rate an fehlenden Abschlusswer-

ten von Patientenseite (hier bis zu 25 %, auch nachdem bei fehlenden Post-Werten auf Katamnese- und Verlaufsmessungen zurückgegriffen wurde) in der weiteren Diagnostik wünschenswert, da sich ein deutlicher Zusammenhang zwischen Misserfolg und fehlenden Abschlusswerten gezeigt hat. Gleichwohl ergab eine Abschätzung, bei der dieser ungünstige Zusammenhang berücksichtigt wurde, keine deutliche Verschiebung in den (Miss-)Erfolgskategorien.

Die von uns eingesetzten Maße decken nicht alle Bereiche möglichen Misserfolgs ab; so wurden etwa im Rahmen der Therapie aufgetretene Probleme im *Umfeld* des Patienten nicht systematisch erfasst. Bei den hier verwendeten Fragebogenmaßen handelt es sich zudem um relativ unspezifische Instrumente. Diese sind weniger änderungssensitiv für eine spezifische psychopathologische Symptomatik als störungsspezifische Maße in den zugehörigen Diagnosegruppen und sollten damit die tatsächliche Veränderung unterschätzen: Störungsspezifische Analysen würden dieser Erwartung nach deutlichere Ergebnisse erbringen (höhere Effektstärken, höhere Raten an Remission und Response, niedrigere Rate an reliablen Verschlechterungen).

Die Verwendung des Reliable Change Index zur Definition klinisch signifikanter Veränderungen ist zwar etabliert, hat aber den Nachteil, dass der Ausgangswert nicht berücksichtigt wird und somit auch ein hoch symptomatischer Abschlusswert möglicherweise als „Response“ gewertet werden kann (z. B. eine Veränderung im BDI von 50 auf 41). Wir haben in zusätzlichen Analysen ein weiteres – prozentuales – Kriterium für Remission berechnet (mindestens 50 % Verbesserung im auffälligen Bereich und mindestens 25 % Verbesserung auf Gesamtskala; Hiller & Schindler, 2011; Lecrubier, 2002) und kamen zu weitgehend mit der RCI-Methode vergleichbaren Ergebnissen (hier nicht dargestellt). Es soll aber darauf hingewiesen werden, dass insbesondere im hochsymptomatischen Bereich (z. B. bei der Analyse spezifischer Outcome-Maße bei speziellen Störungen) eine solche prozentuale Remissions-Bestimmung als konservative Ergänzung genutzt werden sollte.

Manche Einschränkung des hier dargestellten Vorgehens ist auch darin begründet, dass für bestimmte Konstrukte keine klaren Konventionen der Operationalisierung existierten (z. B. für „hoch symptomatischer Abschlusswert“ oder „Therapieerfahrung“), so dass wir ad hoc entsprechende Festlegungen treffen mussten (hier: 1 *SD* über durchschnittlichem Prä-Wert, oder Einteilung der Therapeuten in „innerhalb der ersten Hälfte der Behandlungsphase ihrer Ausbildung“ vs. „in der zweiten Hälfte“ vs. „approbirt“). Auch lagen Vorschläge zur Klassifizierung von Therapieabbrüchen in qualitätsrelevant vs. qualitätsneutral wie die von Hiller et al. (2011) bzw. Cinkaya et al. (2011) zum Zeitpunkt unserer Analysen noch nicht vor, wobei zukünftig hierfür eine einheitliche Operationalisierung angestrebt werden sollte, um die Vergleichbarkeit zu erleichtern.

Misserfolgsvorschung: Implikationen

Die Verschiebung des Fokus‘ von Therapieerfolg auf Therapiemisserfolg hat forschungsmethodische und inhaltliche Besonderheiten, die im vorliegenden Beitrag nur zum Teil aufgegriffen werden können (zu vertiefenden Aspekten siehe Barbrack, 1985; Barlow, 2010; Jacobi, 2001; Kohls, Bents & Pietrowsky, 2002; Lambert, 2011; Lilienfeld, 2007; Margraf, 2009; Mays & Franks, 1985; Mohr, 1995; Sponzel, 2002). Grundsätzlich muss die weitere Analyse und Interpretation solcher Ergebnisverteilungen, wie sie hier für ein repräsentatives ambulantes Setting beschrieben wurden, daraufhin ausgerichtet sein, möglichst den Varianzanteil zu isolieren, der spezifisch auf die Psychotherapie zurückgeht: Sind Therapieerfolg oder Misserfolg „verdient“ oder „unverdient“ (Caspar & Kächele, 2008; Helbig-Lang, 2009; Höfler, Gloster & Hoyer, 2010; Lilienfeld, 2007; Mohr, 1995), d. h. inwiefern haben auch Patientencharakteristika oder andere, z. B. äußere Einflüsse zum Therapieergebnis beigetragen? Dies betrifft die intensiv debattierte Frage, ob Therapeuten und ihr therapeutisches Vorgehen Verschlechterungen *verursacht* oder lediglich *nicht verhindert* haben (Barlow & Nock, 2009; Bienenstein & Rother, 2009; Lutz, 2002). In diesem Zusammenhang sollte auch unser Befund, dass Misserfolge mit einem Trend zu längerer Therapiedauer assoziiert waren, weiter untersucht werden. Wir gehen nicht davon aus, dass gerade diese längere Therapiedauer zu Verschlechterungen oder hoch symptomatischen Abschlusswerten führte, halten es aber dennoch für angezeigt, gerade bei unbefriedigenden (Zwischen-)Ergebnissen die eingesetzte Therapiedosis kritisch zu prüfen.

Die Möglichkeiten, solches zu beforschen, sind im Rahmen einer Standardevaluation wie der vorliegenden begrenzt, weswegen wir dafür plädieren, hierzu ergänzend auch qualitatives Vorgehen einzusetzen (z. B. systematische Befragung von im Rahmen einer solchen Evaluation identifizierten Misserfolgs-Patienten). Grundsätzlich ist dies ein Feld, in dem Einzelfallforschung komplementär zur unseren auf Gruppenvergleichen basierenden standardisierten Ansätzen nützlich bzw. sogar notwendig ist. Die Operationalisierung und Erfassung von Misserfolg entspricht im Übrigen noch nicht den in der Psychotherapieforschung sonst üblichen Standards. So gibt es trotz der umfangreichen Regelungen und Vorgaben für Assessment und Datenauswertung in kontrollierten Studien (RCTs) noch kein etabliertes Vorgehen, Fehlentwicklungen und Therapieschaden, Risiken und Nebenwirkungen oder Behandlungsfehler etc. angemessen einzubeziehen. Ein solches Vorgehen legt jedoch auch eine methodische Schwerpunktverlagerung in Richtung einer stärkeren Berücksichtigung ideografischer gegenüber rein nomothetischer Methoden nahe (Barlow & Nock, 2009).

Die dargestellte Bedeutung der Verteilung der individuellen Werte im Therapieverlauf impliziert, dass der Präsentation von klinisch signifikanten Veränderungen eine wichtige Rolle zukommt. Die hier vorgenommene Darstellung der grafischen Abbildung der Veränderung von An-

fangs- zu Abschlusswerten und der Bericht der Häufigkeit klinisch bedeutsamer (Miss-) Erfolgskategorien sind deutlich informationshaltiger als die Angabe von Effektstärken. Deswegen empfehlen wir, bei der Therapieevaluation solche Analysen einzubeziehen (vgl. Hiller et al., 2011).

Auch wenn die Ergebnisse durchaus ermutigend ausfallen, können wir uns als Psychotherapeuten nicht darauf ausruhen: „Failures are a challenge“ (Emmelkamp & Foa, 1983, S. 1), und Fehlentwicklungen in einzelnen, konkreten Psychotherapien wird es immer geben. Wir können aber kontinuierlich daran arbeiten, solche Fehlentwicklungen zu vermeiden oder zu minimieren, z. B. durch systematische Rückmeldeprozeduren, wie sie von Lambert und Kollegen vorangetrieben werden (Lambert, Whipple, Vermeersch, Smart, Hawkins, Nielsen et al., 2002; Lutz, 2002; Lutz et al., 2011). Auch in Aus- und Fortbildung (Boisvert & Faust, 2006; Castonguay et al., 2010) sowie in der Supervision kann das Ideal des „reflektierenden Praktikers“ gefördert werden, indem eine systematische Auseinandersetzung mit therapeutischem Misserfolg stattfindet.

Literatur

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders – DSM-IV* (4th edition, text revision ed.). Washington: American Psychiatric Association.
- Barbrack, C. R. (1985). Negative outcome in behavior therapy. In D. T. Mays & C. M. Franks (Eds.), *Negative outcome in psychotherapy – and what to do about it* (pp. 76–105). New York: Springer Publishing.
- Barlow, D. H. (2004). Psychological treatments. *American Psychologist*, 59 (9), 869–878.
- Barlow, D. H. (2010). Negative effects from psychological treatments a perspective. *American Psychologist*, 65 (1), 13–20.
- Barlow, D. H. & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science*, 4 (1), 19–21.
- Bergin, A. E. (1963). The empirical emphasis in psychotherapy: A symposium. The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology*, 10 (3), 244–250.
- Bienenstein, S. & Rother, M. (2009). *Fehler in der Psychotherapie: Theorie, Beispiele und Lösungsansätze für die Praxis*. Wien: Springer.
- Boisvert, C. M. & Faust, D. (2006). Practicing psychologists' knowledge of general psychotherapy research findings: Implications for science-practice relations. *Professional Psychology: Research and Practice*, 37 (6), 708–716.
- Caspar, F. & Kächele, H. (2008). Fehlentwicklungen in der Therapie. In C. Mundt, F. Caspar & S. Herpertz (Hrsg.), *Störungsorientierte Therapie* (S. 729–743). München: Urban & Fischer.
- Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R. & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist*, 65 (1), 34–49.
- Cinkaya, F., Schindler, A. & Hiller, W. (2011). Wenn Therapien vorzeitig scheitern. Merkmale und Risikofaktoren von Abbrüchen in der ambulanten Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40, 224–234.
- Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, 59 (7), 614–625.
- Emmelkamp, P. M. G. & Foa, E. B. (1983). Failures are a challenge. In E. B. Foa & P. M. G. Emmelkamp (Eds.), *Failures in behavior therapy* (pp. 1–9). New York: Wiley.
- Fäh, M. (2002). Wenn Analyse krank macht. Methodenspezifische Nebenwirkungen psychoanalytischer Therapien. In M. Märtens & H. Petzold (Hrsg.), *Therapieschäden* (S. 109–147). Mainz: Grünewald.
- Fischer-Klepsch, M., Münchau, N. & Hand, I. (2009). Misserfolge in der Verhaltenstherapie. In J. Margraf & S. Schneider (Hrsg.), *Lehrbuch der Verhaltenstherapie. Band 1: Grundlagen, Diagnostik, Verfahren, Rahmenbedingungen* (3., vollständig bearbeitet und erweiterte Aufl., S. 261–275). Heidelberg: Springer.
- Foa, E. B. & Emmelkamp, P. M. G. (1983). *Failures in behavior therapy*. New York: Wiley.
- Franke, G. H. (2000). *BSI. Brief Symptom Inventory – Deutsche Version. Manual*. Göttingen: Beltz.
- Hansche, Y. (2008). *Inanspruchnahmeverhalten und Therapieabschluss in einer verhaltenstherapeutischen Hochschulambulanz*. Unveröffentlichte Diplomarbeit, TU Dresden.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1995). *Beck-Depressions-Inventar (BDI)* (2ed.). Bern: Huber.
- Helbig-Lang, S. (2009). *Veränderungsrelevante Prozesse jenseits psychotherapeutischer Sitzungen – Konzeption, Beschreibung und Ansätze für ein optimiertes therapeutisches Vorgehen*. Dissertationsschrift, Technische Universität Dresden, Dresden.
- Hiller, W., Bleichhardt, G., Schindler, A. (2009). Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 57, 7–22.
- Hiller, W. & Schindler, A. (2011). Response und Remission in der Psychotherapieforschung. *Psychotherapie – Psychosomatik – Medizinische Psychologie*, 61, 170–176.
- Hiller, W., Schindler, A., Andor, T. & Rist, F. (2011). Vorschläge zur Evaluation regulärer Psychotherapien an Hochschulambulanzen im Sinne einer Phase-IV-Therapieforschung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40, 22–32.
- Hoffmann, S. O., Rudolf, G. & Strauss, B. (2008). Unerwünschte und schädliche Wirkungen von Psychotherapie. Eine Übersicht mit dem Entwurf eines eigenen Modells. *Psychotherapeut*, 53, 4–16.
- Höfler, M., Gloster, A. T. & Hoyer, J. (2010). Causal effects in psychotherapy: Counterfactuals counteract overgeneralization. *Psychotherapy Research*, 20 (6), 668–679.
- Hoyer, J., Becker, E., Neumer, S., Soeder, U. & Margraf, J. (2002). Screening for anxiety in an epidemiological sample: predictive accuracy of questionnaires. *Journal of Anxiety Disorders*, 16, 113–134.
- Hoyer, J., Ruhl, U., Scholz, D. & Wittchen, H. U. (2006). Patients' feedback after computer-assisted diagnostic interviews for mental disorders. *Psychotherapy Research*, 16 (3), 357–363.
- Hoyer, J. & Uhlmann, S. (2008). Therapieevaluation. *Klinische Diagnostik und Evaluation*, 1, 84–106.
- Hoyer, J., Uhlmann, S., Rambow, J. & Jacobi, F. (2009). Reduction of sexual dysfunction: by-product of cognitive-behavioural therapy for psychological disorders? *Sexual and Relationship Therapy*, 24 (1), 64–73.
- Jacobi, F. (2001). Misserfolgsvorschau in der Verhaltenstherapie. In R. Dohrenbusch & F. Kaspers (Hrsg.), *Fortschritte der Klinischen Psychologie und Verhaltensmedizin* (S. 323–346). Lengerich: Pabst.
- Jacobi, F. (2006). Entwicklung und Beurteilung therapeutischer Interventionen. In H.-U. Wittchen & J. Hoyer (Hrsg.), *Klinische Psychologie und Psychotherapie* (S. 553–582). Berlin: Springer.
- Jacobson, N. S. & Truax, P. (1991). Clinical significance – a statistical approach to defining meaningful change in psy-

- chotherapy research. *Journal of Consulting and Clinical Psychology*, 59 (1), 12–19.
- Kohls, S., Bents, H. & Pietrowsky, R. (2002). Prädiktoren für Misserfolg in der verhaltenstherapeutischen Behandlung von Zwangspatienten. *Verhaltenstherapie*, 12, 98–106.
- Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioral Practice*, 18 (3), 413–420.
- Lambert, M. J., Hawkins, E. J. & Hafield, D. R. (2002). Empirische Forschung über negative Effekte in der Psychotherapie: Befunde und ihre Bedeutung für Praxis und Forschung. In M. Märtens & H. Petzold (Hrsg.), *Therapieschäden. Risiken und Nebenwirkungen von Psychotherapien* (S. 40–59). Mainz: Grünewald.
- Lambert, M. J. & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change* (5 ed., pp. 139–193). New York: Wiley & Sons.
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L. et al. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology & Psychotherapy*, 9 (2), 91–103.
- Lecrubier, Y. (2002). How do you define remission? *Acta Psychiatrica Scandinavica*, 106, 7–11.
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2 (1), 53–70.
- Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., Von Wittleben, I., Schroeder, B. et al. (2005). Who comes, who stays, who profits? Predicting refusal, dropout, success, and relapse in a short intervention for social phobia. *Psychotherapy Research*, 15 (3), 210–225.
- Lunnen, K. M., Ogles, B. M. & Pappas, L. N. (2008). A multiperspective comparison of satisfaction, symptomatic change, perceived change, and end-point functioning. *Professional Psychology-Research and Practice*, 39 (2), 145–152.
- Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research*, 12 (3), 251–272.
- Lutz, W., Böhnke, J. R., Köck, K. & Bittermann, A. (2011). Diagnostik und psychometrische Verlaufsrückmeldungen im Rahmen eines Modellprojektes zur Qualitätssicherung in der ambulanten Psychotherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40, 283–297.
- Margraf, J. (2009). Risiken und Nebenwirkungen. In J. Margraf & S. Schneider (Hrsg.), *Lehrbuch der Verhaltenstherapie, Band 1: Grundlagen, Diagnostik, Verfahren, Rahmenbedingungen* (3., vollständig bearbeitete und erweiterte Auflage ed., S. 277–296). Heidelberg: Springer.
- Margraf, J. & Lieb, R. (1995). Was ist Verhaltenstherapie? Versuch einer zukunftsorientierten Neucharakterisierung. *Zeitschrift für Klinische Psychologie*, 24, 1–7.
- Märtens, M. & Petzold, H. (2002). *Therapieschäden*. Mainz: Grünewald.
- Mays, D. T. & Franks, C. M. (1985). *Negative outcome in psychotherapy – and what to do about it*. New York: Springer.
- Mohr, D. C. (1995). Negative outcome in psychotherapy – a critical review. *Clinical Psychology: Science and Practice*, 2 (1), 1–27.
- Noll-Hussong, M. (2011). Primum non nocere, secundus opinio vulnere – Psychotherapie im Fokus des klinischen Risikomanagements. *Psychotherapeut*, 1–5.
- Ogles, B. M., Lunnen, K. M. & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, 21 (3), 421–446.
- Sponsel, R. (2002). Über potenzielle Kunst- oder Behandlungsfehler in der Psychotherapie aus allgemeiner und integrativer Sicht. *Vortrag auf der Ersten Fachtagung des IVS, 27.7.2002.*, Verfügbar unter: www.sgipt.org/gipt/kf/kf020727.htm.
- Steffanowski, A., Kramer, D., Fembacher, A., Glahn, E. M., Bruckmayer, E., Heymann, F. v., Kriz, D., Pfaffinger, I. & Wittmann, W. W. (2011). Praxisübergreifende Dokumentation der Ergebnisqualität ambulanter Psychotherapie in Bayern. Erste Ergebnisse des Modellvorhabens „QS-PSY-BAY“. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 40, 267–282.
- Tress, W. & Erny, N. (2008). Ethik in der Psychotherapie. *Psychotherapeut*, 53 (5), 328–337.
- Wierzbicki, M. & Pekarik, G. (1993). A metaanalysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24 (2), 190–195.
- Wirth, H. J. (2005). Gurutum und Machtmissbrauch in der Psychotherapie. *PiD – Psychotherapie im Dialog*, 6 (2), 136–140.
- Wittchen, H.-U. & Pfister, H. (1997). DIA-X Interview. Frankfurt: Swets and Zeitlinger.
- Wittmann, W. W., Lutz, W., Steffanowski, A., Kriz, D., Glahn, E. M., Völkle, M. C. et al. (2011). *Qualitätsmonitoring in der ambulanten Psychotherapie: Modellprojekt der Techniker Krankenkasse – Abschlussbericht*. Hamburg: Techniker Krankenkasse.

Prof. Dr. Frank Jacobi

Psychologische Hochschule Berlin (PHB)
 Am Köllnischen Park 2
 10179 Berlin
 E-Mail: f.jacobi@psychologische-hochschule.de