

Im Verbund mit den in Kapitel 28.1 skizzierten Kriterien in Bezug auf die *Theorie*, sind damit in Deutschland recht klare Vorgaben formuliert, unter welchen Bedingungen die Anerkennung als wissenschaftliches Psychotherapieverfahren erteilt wird.

38.1.2 Messung der Veränderungen

Für die Beurteilung der Wirksamkeit von Psychotherapie bzw. für Wirksamkeitsvergleiche zwischen verschiedenen Therapieverfahren ist entscheidend, dass die Beurteilung anhand von möglichst *konsensfähigen* Veränderungskriterien erfolgt, dass es also einen Konsens darüber gibt, mit welchen Messungen zwischen »erfolgreichen« und »nicht erfolgreichen« Behandlungen unterschieden werden kann. »Ein solcher Konsens existiert allerdings nicht« (Margraf 2009). Die Messung von Therapieerfolg ist immer noch äußerst vielfältig: Froyd et al. (1996) beispielsweise fanden in 348 Psychotherapiestudien (publiziert zwischen 1983 und 1988) 1430 verschiedene Maße für den Therapieerfolg; und Lambert & Hill (1994) zeigen an einem Beispiel, dass die »Erfolgsquote« in ein und derselben Studie in Abhängigkeit von der Wahl verschiedener Erfolgsindikatoren zwischen 6,39 % und 69 % schwanken kann.

Abgesehen von der Problematik der Wahl des angemessenen Erfolgsmaßes, bestehen in verschiedenen Therapieschulen prinzipielle Vorbehalte gegenüber quantifizierenden Messungen von durch Psychotherapie erreichten Veränderungen. So argumentieren Psychoanalytiker (aber auch Vertreter anderer Schulen) häufig, dass sich die »eentlichen« Veränderungen im Bereich unbewusster Strukturen vollziehen, die sich nun einmal nicht quantifizieren lassen; Veränderungsprozesse seien zudem derart individuell, dass quantifizierende Vergleiche mittels standardisierter Maße sinnlos seien;

die Konzentration auf die Veränderung von Beschwerden und Symptomen sei zu oberflächlich, um die tatsächlichen Veränderungen abzubilden; Psychotherapie sei weit mehr, als eine Methode zur Reduzierung von Symptomen, sodass Symptomveränderungen alleine wenig über den Erfolg eines therapeutischen Prozesses aussagen etc.

Diese Einwände sind zwar nicht von der Hand zu weisen – daraus aber eine generelle Ablehnung gegenüber empirischen Studien zur Dokumentation von (auch symptomatikbezogenen) Veränderungen abzuleiten, erscheint verfehlt und insbesondere gegenüber den Patienten unethisch, die Hilfe suchen. Patienten kommen üblicherweise aufgrund ihrer psychischen Störungen zum Psychotherapeuten, mit dem klaren (und nachvollziehbaren) Anliegen, dass die ihm angebotene Therapie zur Heilung oder zumindest Linderung eben dieser Beschwerden beiträgt. Die Frage eines Patienten, ob die Wirksamkeit der vom Therapeuten praktizierten Therapie für den Beschwerdebereich des Patienten belegt ist, erscheint vollkommen legitim. Das gilt auch für Therapieformen, deren Veränderungsstrategien wenig bis gar nicht an der manifesten Symptomatik ansetzen. Letztlich gilt es auch hier zu zeigen, dass Veränderungen in den vom jeweiligen Verfahren als relevant angesehenen Bereichen mit Veränderungen auf der Ebene der manifesten Beschwerden einhergehen. Gleichgültig an welcher psychischen Dimension psychotherapeutisch primär gearbeitet wird, am Ende sollte sich dies immer auch auf der Ebene der subjektiven Beschwerden positiv auswirken. Ein Psychotherapieverfahren, welches psychische Störung nicht auch auf der Ebene der leidvollen Symptomatik positiv beeinflussen kann, ist kein Psychotherapieverfahren. Es ist selbstverständlich richtig, dass jede durchgeführte Behandlung sich auf den individuellen und für sich einzigartigen Menschen neu einstellen muss, dass damit jeder individuelle Therapieprozess ein einzigartiges Gesche-

hen darstellt, dass therapeutische Prozesse daher im Einzelfall auch scheitern können, dass es im Einzelfall auch vorkommt und dass Patienten von einer Therapie persönlich profitieren können, obwohl die Symptomatik weiterbesteht. Dennoch sollte jedes psychotherapeutische Verfahren konsensfähige Belege erbringen, dass die Durchführung dieses Verfahrens mit einer hinreichend hohen Wahrscheinlichkeit zur Heilung oder zumindest Linderung derjenigen Störungen und Beschwerden führen kann, derentwegen die Therapie begonnen wurde.

Schulte (1993; vgl. Margraf 2009) schlägt einen allgemeinen **Konsens zur Erfolgsmessung** vor: Da das (gesellschaftlich vorgegebene) Ziel einer Therapie in der Heilung bzw. Linderung von Störungen mit Krankheitswert liegt, sollte »Therapieerfolg« (u. a.) in einem Rückgang der Beschwerden und ihrer Folgen bestehen. Deshalb sollte die Erfassung des Behandlungserfolgs immer auch **störungsspezifische Messungen** beinhalten, die sich an den aktuellen Diagnoseschlüsseln (DSM, ICD) orientieren. Darüber hinaus sollten auch weitere **Störungsfolgen** (Demoralisation, Krankenrolle, Verlust von Kontrolle, Selbstwert und sozialer Einbindung), die störungsübergreifende Symptombelastung und mögliche Nebenwirkungen der Therapie erhoben werden.

Sich allerdings nur auf Selbsteinschätzungen der Patienten zu beschränken, kann auch in die Irre führen, wie verschiedene Studien zeigen (z.B. Shedler et al. 1993; ► **Kasten**).

Shedler et al. (1993) untersuchten »**The illusion of mental health**«: Versuchspersonen füllten gängige Fragebögen zu ihrem psychischen Befinden aus (Neuroticism Scale von Eysenck bzw. Beck-Depression-Inventar), und wurden zudem mit dem Early Memory Test (EMT – ein »free response measure«) untersucht. Versuchspersonen, die in den Selbstangaben

per Fragebogen »gesund« erschienen aber »ungesund« laut Ratings ihrer EMT-Aussagen zu sein schienen, zeigten stärkere Herzaktivität unter Stressbedingungen, als Versuchspersonen, die in beiden Verfahren als gesund eingestuft wurden. Die »illusory mental health« Personen zeigten sogar eine stärkere Herzaktivität unter Stress, als diejenigen, die in beiden Verfahren (Selbstangaben und EMT) als nicht gesund eingestuft wurden!

Scher et al. (2005) geben einen Überblick über den Einsatz **experimenteller Methoden** zur Messung von Therapieerfolg in kognitiv-behavioralen Therapiestudien; dabei geht es meist um »Priming-Verfahren« zur Aktivierung von latenten Schemata und der Erfassung von indirekten (»impliziten«) Priming-Effekten zur Erfolgsmessung. Hedlund & Rude (1995) verglichen beispielsweise symptomatisch erfolgreich behandelte depressive Patienten mit Versuchspersonen, die nie depressiv waren: In den Fragebogen-Scores unterschieden sich beide Gruppen nicht voneinander in Bezug auf »dysfunctional thinking«, aber nach einer »self-focus manipulation« zeigten die ehemals depressiven Patienten deutlich erhöhte Werte für negative »information processing«; dies war bei den nie-depressiven Versuchspersonen nicht der Fall. Hier zeigt sich also, dass Psychotherapie zu symptomatischer Remission und selbstberichteter Verbesserung führen kann, ohne notwendigerweise die zugrundeliegende potenziell pathogene Struktur zu verändern.

Segal et al. (1999) verglichen zwei Gruppen von Patienten: Die eine war symptomatisch erfolgreich mit KVT behandelt worden, die andere mit Pharmako-Therapie. Bei einer »negative mood induction« zeigten die pharmakologisch behandelten Patienten deutlich mehr dysfunktionale Kognitionen als die mit KVT behandelten. Zudem war

die Reaktion auf die mood induction prädictiv für die Rückfälle im Katamnesezeitraum. Dies zeigt einerseits, dass die Psychotherapie offenbar andere und tiefergehende Veränderungen erzeugte als die Pharmakotherapie und andererseits, dass diese tiefergehenden Veränderungen ein Schutz vor weiterbestehenden Vulnerabilitäten darstellen. Zudem wird deutlich, dass diese *Vulnerabilität in Form von »ruhenden« Schemata* am besten unter Bedingungen getestet wird, die geeignet sind, diese Schemata zu aktivieren (was bei der üblichen Symptomabfrage nicht der Fall ist).

Josephs und Bornstein (2011) ziehen aus diesen Untersuchungen die Schlussfolgerung, dass zu jeder guten Messbatterie möglichst auch »implizite« Messungen (unter Stressbedingungen) gehören sollten, gerade auch zur Erfassung der in psychoanalyti-

schen Therapien intendierten »Umstrukturierung«.

Innerhalb der OPD-Arbeitsgruppe wurde die Heidelberger **Umstrukturierungsskala** (HSCS: Heidelberg Structural Change Scale; Rudolf et al. 2000; Arbeitskreis OPD 2006) entwickelt, die das Niveau der innerpsychischen Veränderung jenseits der Symptomatik erfasst. Bei der HSCS werden zu Behandlungsbeginn auf der Basis der OPD-Diagnostik *Foki* festgelegt (Beziehungsmuster, Konfliktthemen, Strukturelle Probleme). Anhand von späteren OPD-Interviews (z.B. bei Behandlungsende) wird dann eingeschätzt, wo der Patient in Hinblick auf die Veränderung dieser spezifischen Foki steht. Die Skala der HSCS ist hier in Kurzform wiedergegeben (► Abb. 38.1).

Die ersten Anwendungen scheinen vielversprechend: Ein höheres Ausmaß der zum

1. Nichtwahrnehmung des Fokusproblems	1 1+	Völlige Abwehr bzw. Vermeidung des Fokusbereichs, es gibt „kein Problem“	
2. Ungewollte Beschäftigung mit dem Fokus	2- 2 2+	Symptomdruck, interpersonelle Schwierigkeiten: Zumutungen, von außen kommend erlebt	
3. Vage Fokuswahrnehmung	3- 3 3+	Passive Beschäftigung mit dem Fokus, ansatzweise Anerkennung, Ahnung eigener Verantwortung	Bewältigung
4. Anerkennung und Erkennung des Fokus	4- 4 4+	Interessiertes Problemverstehen, Arbeitsbeziehung, aktive „Bewältigung“, Handeln	
5. Auflösung alter Strukturen im Fokusbereich	5- 5 5+	Abwehr wird brüchig, Prozess wird zur „Passion“, Trauer, Ausgeliefertsein, Verwirrung	Strukturelle Veränderung
6. Neustrukturierung im Fokusbereich	6- 6 6+	Versöhnliches Erleben, neue Erlebens-/Verhaltensmöglichkeiten stellen sich spontan ein	
7. Auflösung des Fokus	7- 7	Integration, Selbstübereinstimmung, realitätsgerechtes Erleben, Neugestaltungen	

Abb. 38.1: Heidelberger Umstrukturierungsskala (Arbeitskreis OPD 2006).

Behandlungsende erreichten Umstrukturierung (in Analytischen und Tiefenpsychologischen Psychotherapien) hing signifikant mit einer stärker ausgeprägten Symptomreduzierung (SCL-90), mit weniger interpersonellen Problemen (IIP) sowie mit größerer Behandlungszufriedenheit von Patienten und Therapeuten zusammen. Von den zum Behandlungsende erfassten Veränderungsmaßen (Symptomreduktion, Veränderung der interpersonellen Probleme, Umstrukturi-

erung) erlaubte nur die *Umstrukturierung* eine Vorhersage der Lebenszufriedenheit drei Jahre nach Therapieende (Grande et al. 2009; Rudolf et al. 2012).

Die Veränderungseffekte einer therapeutischen Maßnahme werden üblicherweise in Form von sogenannten **Effektstärken** dargestellt. Dabei handelt es sich um ein statistisches Maß zur Quantifizierung der Größe eines beobachteten Effektes bzw. Unterschiedes.

Grundsätzlich beziehen sich Effektstärken (ES) meist auf zwei Vergleiche:

- Effektstärke der Veränderung *innerhalb einer Gruppe über die Zeit* (z.B. Störungsschwere vor Behandlungsbeginn im Vergleich zur Störungsschwere zum Behandlungsende oder zum Katamnese Zeitpunkt).
- Effektstärke *zwischen zwei Behandlungsgruppen bzw. zwischen Behandlungsgruppe und unbehandelter Kontrollgruppe* (z.B. Vergleich der Störungsschwere der mit zwei verschiedenen Psychotherapieverfahren behandelten Patientengruppen am Behandlungsende oder zum Katamnese Zeitpunkt).

Die Effektstärke wird als standardisierter Differenzwert berechnet, wie in der folgenden Formel ausgedrückt (Hedges 1981):

$$ES = \frac{\text{Mittelwert Gruppe 1} - \text{Mittelwert Gruppe 2}}{\text{gepoolte Standardabweichung innerhalb der Gruppen}}$$

Eine Effektstärke von 1.0 beispielsweise bei einem prä-post-Vergleich innerhalb einer Gruppe besagt, dass der durchschnittliche Patient dieser Gruppe am Behandlungsende einen Wert im betrachteten Zielkriterium aufweist, der genau eine Standardabweichung über derjenigen des durchschnittlichen Patienten vor Behandlungsbeginn liegt. Bei einer Effektstärke von 0.8 liegen die prä-post-Durchschnitte 0.8 Standardabweichungen auseinander etc.

Bei einem Vergleich zwischen zwei Treatmentgruppen besagt eine Effektstärke von 1.0 entsprechend, dass der durchschnittliche Patient aus Treatment-Gruppe 1 (z.B. mit Psychotherapie behandelte Patienten) einen Wert im betrachteten Zielkri-

terium aufweist, der genau eine Standardabweichung über derjenigen des durchschnittlichen Patienten aus Gruppe 2 (z.B. ohne Psychotherapie) liegt.

Als Konvention zur Interpretation von Effektstärken (ES) haben sich folgende Zuordnungen eingebürgert:

Effektstärke (ES)	Interpretation
ES > 0.80	große Effekte
ES 0.50–0.80	mittlere Effekte
ES 0.20–0.50	kleine Effekte
ES < 0.20	Effekte ohne praktische Bedeutsamkeit

Bei *prä-post*-Vergleichen bzw. *prä-Katamnese*-Vergleichen innerhalb einer mit einem Psychotherapieverfahren behandelten Patientengruppe sind hohe ES zu erwarten und diese werden üblicherweise auch erreicht (► **Kap. 38.3.1**). Auch bei *post*-Vergleichen zwischen mit Psychotherapie behandelten Patienten und Patienten in Wartegruppen (oder anderen Kontrollbedingungen) werden üblicherweise mittlere bis hohe Effektstärken erreicht.

Bei Vergleichen *zwischen zwei Psychotherapieverfahren* sind keine großen Effektstärken zu erwarten (es sei denn, man ginge von der Unwirksamkeit eines der

Verfahren aus). Um evtl. dennoch vorhandene statistische Unterschiede z. B. am Behandlungsende oder nach längeren Katamnese Zeiträumen zu finden, sind relativ große Stichproben notwendig (Stichwort: »Power« eines statistischen Vergleichstests).

Neben der statistischen Signifikanz von Veränderungen ist die sogenannte **klinische Signifikanz** von Bedeutung. Allerdings liegen hier keine einheitlichen Standards zur Bestimmung vor. Eine häufig angewandte Berechnung der klinischen Signifikanz ist die von (Jacobson & Truax 1991).

Klinische Signifikanz nach Jacobson und Truax (1991):

Eine *klinisch signifikante Symptomverbesserung* liegt demnach vor, wenn

- a. sich die Symptomatik eines Patienten nach dem *Reliable Change Index* (RCI)¹⁷⁷ statistisch signifikant verbessert hat ($RCI > 1.96$) und
- b. die Symptomatik bei Therapieende innerhalb eines funktionalen Normbereichs liegt.

Eine *klinisch signifikante Symptomverschlechterung* liegt vor, wenn

- a. sich die Symptomatik eines Patienten statistisch signifikant verschlechtert hat ($RCI < -1.96$) und
- b. die Symptomatik bei Therapieende innerhalb eines dysfunktionalen Normbereichs liegt.

Die Bestimmung der klinischen Signifikanz erlaubt es beispielsweise, zwei Behandlungsmethoden dahingehend zu vergleichen, mit welcher Methode mehr Patienten eine klinisch signifikante Besserung erreichen (und weniger Verschlechterungen).

38.1.3 Beurteilung der Wirksamkeits- »Evidenz« von Psychotherapie

Für die Bewertung der Wirksamkeit von psychotherapeutischer und psychopharma-

¹⁷⁷ Der RCI ist mit einem z-Wert vergleichbar, d. h. der Wert von 1.96 bedeutet, dass die Differenz zweier Werte knapp zweimal so groß ist wie ihr Standardmessfehler und insofern (mit einer Restfehlerwahrscheinlichkeit von 5 %) die Nullhypothese verworfen werden kann, dass kein Unterschied zwischen beiden Werten besteht.

kologischer Interventionen wird üblicherweise die »Evidenz«-Gradeinteilung für klinische Studien in Anlehnung an Chambless & Hollon (1998), Segal et al. (2001), Ärztliche Zentralstelle Qualitätssicherung (2003) und anderen verwendet.

Chambless und Hollon (1998) definieren »empirically supported treatment« (EST) durch eine Mischung aus unterschiedlichen Studientypen: »It is suggested that, in evaluating the benefits of a given treatment, the

greatest weight should be given to efficacy trials but that these trials should be followed by research on effectiveness in clinical settings and with various populations and by cost-effectiveness research« (Chambless & Hollon 1998, S. 7).

Tabelle 38.2 gibt die Evidenzgradeinteilung wieder, wie sie beispielsweise in der AWMF-Leitlinie Persönlichkeitsstörungen und der DGPs-Leitlinie Panikstörungen/ Agoraphobie angewendet wurde¹⁷⁸.

Tab. 38.2: Evidenzstufen (nach Chambless & Hollon 1998, aus Heinrichs et al. 2009)

»Evidenz«-Grad	»Evidenz«-Basis	Beurteilung
Ia	Metaanalyse(n) über mehrere randomisierte, kontrollierte Studien	(I) wirksam
Ib	Mindestens zwei randomisierte, kontrollierte Studien (RCT) aus unabhängigen Gruppen	
IIa	Eine randomisierte, kontrollierte Studie (RCT)	(II) möglicherweise wirksam
IIb	Serie von gut angelegten quasi-experimentellen Studien (effectiveness Studien, prospektive Kohortenstudien, Fallkontrollstudien, experimentelle Einzelfallstudien)	
III	Nicht-experimentelle oder deskriptive Studien (1-Gruppenprä-post-Vergleiche, Korrelationsstudien)	(III) und (IV) bislang ohne ausreichende Nachweise
IV	Unsystematische Einzelfallstudien, Kasuistiken, Experten, Konsensuskonferenzen, klinische Erfahrung	

Die Fokussierung auf RCTs ist sehr deutlich. Ein Psychotherapieverfahren, für welches beispielsweise zwar eine Reihe von gut angelegten, naturalistischen Wirksamkeitsstudien vorliegt, in denen aber keine randomisierte Patientenzuteilung erfolgte, kann in diesem Schema maximal den Evidenzgrad IIb erreichen und damit maximal als »möglicherweise wirksam« gelten – entsprechend wird ihr Einsatz in den Leitlinien

im besten Fall mit Einschränkungen empfohlen.

Der ursprüngliche Vorschlag von Chambless und Hollon (1998), in die Definition von EST auch Effectiveness- und Kosten-Nutzen-Analysen mit einzubeziehen, wird bisher nicht realisiert: Als *empirically supported* gilt eine Therapie, wenn sie einen Evidenzgrad Ia oder Ib nach obigem Schema erreicht.

178 Siehe: <http://leitlinien.net/> sowie Homepage der DGPs, Fachgruppe Klinische Psychologie (www.klinische-psychologie-psychotherapie.de).