

## 30 Entwicklung und Beurteilung therapeutischer Interventionen

*F. Jacobi*

- 30.1 Perspektiven der Interventionsforschung:  
Das Wechselspiel zwischen Entwicklung und Beurteilung – 610**
  - 30.1.1 Legitimation und Evaluation versus Neuentwicklung von Verfahren – 610
  - 30.1.2 Psychotherapietheorien neuer Generation – 611
- 30.2 Phasen der Therapieevaluation – 613**
  - 30.2.1 Vorbemerkung: Wissenschaftliche Fundierung versus wissenschaftliche Anerkennung – 613
  - 30.2.2 Phasen der Psychotherapieprüfung in Analogie zur Pharmaprüfung – 614
  - 30.2.3 Die besondere Bedeutung der Phase III:  
Anforderungen an Wirksamkeitsstudien – 616
  - 30.2.4 Ausblick: »Patient focused therapy research« als Phase V  
der Therapieevaluation? – 623
- 30.3 Überblick zum Forschungsstand bei verschiedenen  
Interventionsverfahren – 627**
  - 30.3.1 Unterschiedliche Therapieformen – gleiche Wirkung? – 627
  - 30.3.2 Empirically Supported Treatments (EST) – 629
- 30.4 Die Erforschung von Wirkfaktoren und Wirkmechanismen – 631**
  - 30.4.1 »Allgemeine« und »unspezifische« Faktoren:  
Psychologische Intervention = Placebo? – 632
  - 30.4.2 Sollten klinisch-psychologische Interventionen standardisiert oder  
individualisiert durchgeführt werden? – 632
  - 30.4.3 Beispiel für die systematische Weiterentwicklung einer Therapie  
der posttraumatischen Belastungsstörung – 633
  - 30.4.4 Diagnosetechnologien – 635
- 30.5 Die perfekte therapeutische Intervention – 635**
- Literatur – 638**



### 30.1 Perspektiven der Interventionsforschung: Das Wechselspiel zwischen Entwicklung und Beurteilung

Angesichts der Vielzahl vorliegender klinisch-psychologischer Interventionsverfahren (► Kap. 19) stellt sich die Frage nach ihrer jeweiligen wissenschaftlichen Fundierung: Hat die Intervention eine theoretisch begründete Basis? Inwiefern ist ihre Wirksamkeit in kontrollierten Studien nachgewiesen? Welche Erkenntnisse gibt es darüber hinaus für die Effektivität im Versorgungsalltag? Welche Implikationen hat der – z. T. heterogene – empirische Forschungsstand zu einer bestimmten Intervention für zukünftige Forschung und Weiterentwicklungen? Und nicht zuletzt: Wie sollen sich Praktiker im konkreten Einzelfall verhalten, bzw. wie können im Einzelnen angewandte Interventionen möglichst gut – im Sinne einer wissenschaftlichen Fundierung – begründet werden (Westmeyer, 2000)?

Historisch wurden solchen Fragen seit den 1950er Jahren anhand unterschiedlicher Forschungsperspektiven untersucht. Hierbei kann hinsichtlich des »Auflösungsgrades« des Untersuchungsgegenstandes eine Entwicklung von sehr allgemeinen Konzepten (z. B. Wirkt *die* Psychotherapie an sich?) bis hin zu sehr spezifischen Fragestellungen festgestellt werden (z. B. Was muss speziell bei traumatisierten Patienten hinsichtlich der Methode der Reizkonfrontation beachtet werden?).

#### 30.1.1 Legitimation und Evaluation versus Neuentwicklung von Verfahren

Nach Grawe (1997) können folgende (sich historisch z. T. überlappende) Phasen psychologischer Interventionsforschung skizziert werden:

##### Phasen psychologischer Interventionsforschung

1. **Legitimationsphase:** Ist Psychotherapie generell wirksam?
2. **Wettbewerbsphase:** Welche Form der Psychotherapie ist besser (oder sogar am besten)?
3. **Verschreibungsphase:** Welche Form der Psychotherapie ist bei wem (unter welchen Umständen) indiziert?«
4. **Prozessforschungsphase** zur grundlegenden Frage: Auf welche Weise wirkt Psychotherapie?

##### Legitimationsphase

In ihren Anfängen war die Entwicklung psychotherapeutischer Verfahren fest in Therapieschulen (z. B. psychodynamischen, behavioristischen oder humanistischen) verankert. In den 1950er Jahren dominierten im Praxisalltag klassisch psychoanalytische und verwandte Therapieformen, und als empirische Basis dienten meist unsystematische Fallstudien. Vor diesem Hintergrund äußerte Eysenck (1952) anhand der Auswertung von Falldokumentationen stationärer Einrichtungen die provozierende Kritik, Psychotherapie könne – mit gewissen Unterschieden je nach Orientierung – im besten Fall eine spontane Remission nicht verhindern, auf jeden Fall aber keinen über eine spontane Remission hinaus gehenden Beitrag zur Verbesserung psychischen Leidens liefern! Damit galt es also im Sinne der grundsätzlichen Legitimation psychotherapeutischer Verfahren erst einmal zu belegen, dass Psychotherapie *überhaupt* wirkt und einen spontanen Krankheitsverlauf systematisch positiv (verglichen mit Nichtbehandlung) beeinflussen kann.

##### Wettbewerbsphase

Angespornt von den Bemühungen, psychotherapeutische Interventionen an sich zu legitimieren (z. B. in experimentellen Studien) folgte in den folgenden Jahrzehnten ein Wettbewerb der Therapieschulen, z. B. »Was wirkt besser: Psychoanalyse oder Verhaltenstherapie?«. Auch hier war der Auflösungsgrad hinsichtlich Operationalisierung der untersuchten Intervention meist noch sehr grob (z. B. »Verhaltenstherapie« als Obergruppe für viele mögliche Techniken und Strategien), ebenso wie der störungsspezifische Auflösungsgrad (z. B. »XY hilft bei neurotischen Patienten, nicht aber bei psychotischen«).

##### Verschreibungsphase

Spätestens seit den 1990er Jahren und dem Aufkommen der störungsspezifischen Perspektive (»empirically supported treatments«; ► Abschn. 30.3.2) differenzierte sich die Entwicklung und Überprüfung neuer therapeutischer Ansätze zunehmend. Analog zur Medikamentenverschreibung sollen spezifische Interventionen spezifischen Problemen zugeordnet werden. Nun stehen zunehmend indikationsbezogene Programmpakete im Fokus, die sich häufig auch nicht mehr eindeutig verschiedenen Therapieschulen zuordnen lassen. Diese Verfahren sind hinsichtlich ihrer Wirksamkeit geprüft, können prinzipiell aber immer noch unzureichend (oder sogar falsch) theoretisch fundiert sein.

##### Prozessforschungsphase

Die Untersuchung der Prozesse, die der Wirkung klinisch-psychologischer Interventionen zugrunde liegen, hat eine eigene Qualität: Während die bisher genannten drei Per-

spektiven konservativer Natur sind (Legitimation, Nachweis der Überlegenheit von Verfahren, Optimierung von Indikationsentscheidungen, Minimierung von Risiken), hat dieser Ansatz vor allem progressives Potenzial, denn er kann als Grundlage für die Neu- bzw. Weiterentwicklung von Interventionen dienen. Die grundlegende Frage »Auf welche Weise wirkt Psychotherapie?« wurde zwar bereits seit den Anfängen der Entwicklung von Therapieschulen bzw. Theorien gestellt; es handelt sich hierbei also nicht um die chronologisch aktuellste Phase der Therapieforschung, sondern um während allen bereits genannten »konservativen« Phasen parallel unternommene Bemühungen, die gefundenen positiven Effekte zu erklären. Dennoch lassen sich zwei explizit moderne Strömungen einer solchen Prozessforschung ausmachen:

- a) In den letzten Jahren ist ein rasanter Anstieg grundlagenwissenschaftlichen Inputs bei der Entwicklung von Verfahren zu verzeichnen (z. B. aus Gedächtnis- oder Motivationsforschung, psychobiologischer Forschung, »cognitive sciences«).
- b) Es sind Theorien neuerer Generation im Entstehen begriffen, die bisherige Ergebnisse der Psychotherapieprozessforschung neu – d. h. insbesondere unabhängig von Therapieschulen – integrieren.

#### Wichtig

Wir müssen also bei der wissenschaftlichen Fundierung klinisch-psychologischer Interventionen zwischen der »progressiven« Neuentwicklung von Verfahren im Zuge der theoretischen und empirischen Weiterentwicklung von Konzepten und der »konservativen« Bewertung bzw. Evaluation bereits bestehender Verfahren trennen.

Beide Ansätze sollten dabei nicht gegeneinander ausgespielt werden. Auch wenn etwa ein Wirksamkeitsnachweis oder ein Befund zur differenziellen Effektivität eines Verfahrens zunächst erst einmal nichts Neues schafft bzw. nichts verbessert, sondern eher bereits Bestehendes oder soeben neu Entwickeltes bestätigt, ist der »konservative« Ansatz aus zwei Gründen unerlässlich: Erstens müssen neue, anhand aktueller Prozesstheorien entwickelte Verfahren regelmäßig wieder neu getestet werden, um auf dieser Grundlage weiterzuarbeiten oder nötige Modifizierungen vorzunehmen (»Passen die Daten der neuen Intervention immer noch zu meiner Theorie, die der Neuentwicklung zugrunde liegt?«). Zweitens kann die bestehende Versorgungslandschaft durch den legitimierenden Ansatz dahingehend verbessert werden, dass auf wissenschaftlicher Grundlage Entscheidungshilfen bereitgestellt werden, welche Patienten bei welchen Problemen am besten welche Behandlung erhalten sollten.

### 30.1.2 Psychotherapietheorien neuer Generation

Herkömmliche Therapieschulen haben den Nachteil, dass Weiterentwicklung und Legitimation immer nur im Rahmen ihrer theoretischen Grenzen beforcht wurden. Grawe (1995, 1998) weist darauf hin, dass etwa traditionelle tiefenpsychologisch orientierte Therapieforschung die Befunde der traditionellen verhaltenstherapeutischen Forschung nicht zur Kenntnis nimmt – und umgekehrt –, wenn sie nicht ins althergebrachte Konzept passen. So würde etwa, vereinfacht ausgedrückt, ein herkömmlicher Psychoanalytiker im Rahmen der Therapie konkrete Hilfen zur Problembewältigung wahrscheinlich grundsätzlich als kontraproduktiv abwerten und möglicherweise Folgendes behaupten: »Herumdoktorieren an Symptomen fördert die Verdrängung der eigentlichen Problematik und kann deswegen nur schädlich wirken – zudem ist dieser Ansatz intellektuell zu schlicht, um für komplexes neurotisches Verhalten relevant zu sein!«. Ist eine solche Haltung absolut, werden aus der nicht-psychoanalytischen Forschung kommende Befunde ausgeblendet, die belegen, dass es in bestimmten Konstellationen sehr wohl angezeigt ist, konkrete Hilfestellungen zu geben (z. B. Training sozialer Kompetenzen, Aufbau angenehmer Aktivitäten, »supportives« Therapeutenverhalten).

Umgekehrt könnte bei einer allzu traditionellen verhaltenstherapeutischen Auffassung systematisch übersehen werden, dass bei bestimmten Problemkonstellationen nicht Problemlösen oder neue Copingstrategien erlernt werden müssen, sondern motivationale Konflikte therapeutisch bearbeitet werden sollten. Hier würden also Fakten ausgeblendet, die aus der nicht-verhaltenstherapeutischen Forschung hervorgebracht wurden. Ein dogmatischer Behaviorist würde z. B. die Befunde eines Großteils der Forschung zu humanistischen oder bestimmten kognitiven Verfahren gar nicht erst ernsthaft zur Kenntnis nehmen, z. B. nach dem Motto: »Wir arbeiten daran, unangepasstes Verhalten ab- und erwünschtes Verhalten aufzubauen, und nicht daran, den Sinn des Lebens zu finden!«.

Ursprünglich hatten die traditionellen Therapieschulen (»Psychotherapeutische Theorien erster Generation«; Grawe, 1995) eine wichtige Aufgabe, indem sie zuerst einmal den Erfahrungsbereich der Psychotherapie konstituierten und aus ihren jeweiligen theoretischen Blickwinkeln heraus auch eine reichhaltige empirische Faktenbasis geschaffen haben, die vorher schlichtweg noch nicht bestand. Die Funktion der Theorien bestand im Wesentlichen darin, eine neue Perspektive einzunehmen, aus der bestimmte Phänomene und Zusammenhänge erst als solche erkennbar werden. Das Problem besteht aber darin, dass hierbei oft die Feststellung solcher Fakten mit deren Erklärung bzw. Inter-

## Exkurs

**Einige große Überblickswerke**

Im Jahr 1967 starteten Allen E. Bergin und Sol L. Garfield ein groß angelegtes Projekt zur Dokumentation von Psychotherapieforschung: In regelmäßigen Abständen sollte in einem umfangreichen Handbuch der aktuelle Stand der Psychotherapieforschung von ausgewiesenen Experten therapieschulenübergreifend zusammengefasst werden. Mittlerweile liegt von diesem »Handbook of Psychotherapy and Behavior Change« die 5. Auflage vor (Lambert, 2004); die erste Auflage erschien 1971. Als Einstieg empfiehlt sich das letzte Kapitel, in dem im Ausblick noch einmal überblicksartig wichtige Inhalte angesprochen werden. Das auch als »The Bible« titulierte Buch eröffnet einen Einblick in die Komplexität, die dieser Forschungsbereich mittlerweile erlangt hat.

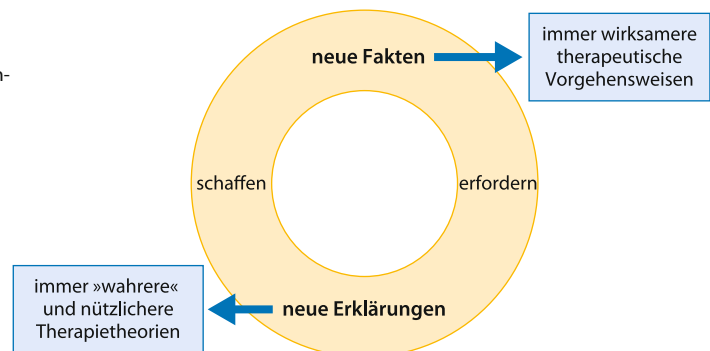
Im deutschsprachigen Raum ist die Integration von Ergebnissen der Psychotherapieforschung in besonderem Maße mit dem Namen Klaus Grawe verbunden. Zunächst lieferte er mit seiner Berner Arbeitsgruppe im Zuge einer bis dato einzigartigen Sammlung von Befunden (Grawe,

Donati & Bernauer, 1994) wichtige Impulse für eine neue Auseinandersetzung mit wissenschaftlichen Ergebnissen der Psychotherapieforschung. In seinem Buch »Psychologische Therapie« (1998) wird die Vielzahl an Ergebnissen zu Wirksamkeit und vor allem zur Wirkungsweise psychotherapeutischen Geschehens in einer Wirkfaktorentheorie integriert, zu deren Kernpunkt die Bearbeitung von »Inkonsistenz«, z. B. aufgrund widersprüchlicher motivationaler Ziele oder aufgrund nicht realisierter Grundbedürfnisse, gehört. Er versucht dabei, in der Form eines Dialogs zwischen einer Psychotherapeutin, einem psychologischen Grundlagenforscher und einem Therapieforscher auf der Basis dieser Befunde eine allgemeine Therapietheorie zu entwickeln. Eine Erweiterung seines Ansatzes, in dem in besonderem Maße das Eingebettetein psychologischer Mechanismen in biologische Strukturen thematisiert wird, lieferte Grawe (2004) mit seinem letzten großen Werk »Neuropsychotherapie«. Mit seinem Tod 2005 verlor die deutschsprachige Klinische Psychologie einen ihrer wichtigsten Vertreter.

pretation vermischt wurde – wobei die Erklärung der gesamten Faktenlage allerdings keiner »Theorie erster Generation« befriedigend gelingen konnte, da in der Regel die Fakten irgendwann nicht mehr in den theoretischen Rahmen passten. Grawe verweist hier auf eines der wichtigsten wissenschaftlichen Grundprinzipien, nämlich die Pflicht zur Rezeption *aller* Fakten, die zum Aussagebereich der jeweiligen Theorie gehören: Es gibt keine verhaltenstherapeutischen oder psychoanalytischen Fakten, sondern wir können inzwischen auf eine äußerst umfangreiche objektive Befundlage zurückgreifen, deren Existenz mittlerweile in weiten Teilen von den ursprünglichen Theorien unabhängig geworden ist. Einen Einblick in diesen aus verschiedensten Perspektiven angehäuften Faktenberg gibt das seit 1971 regelmäßig neu überarbeitete »Handbook of Psychotherapy and Behavior Change« (► Exkurs).

Als Lösung dieses Problems der zunehmenden Unmöglichkeit, das ganze Spektrum der Befunde mit den »Theorien erster Generation« erklären zu können, schlägt Grawe (1995) aus der neusten empirischen Datenbasis heraus »post hoc« entwickelte »Theorien zweiter Generation« vor. In dem Maße, in dem solche Theorien oder Modelle die Gesamtheit der Fakten besser erklären können als die Ursprungstheorien, verlieren die Ursprungstheorien an Nützlichkeit und Aktualität. ■ Abb. 30.1 veranschaulicht diesen gewissermaßen dialektischen, spiralförmigen Prozess: Immer neue Fakten schaffen die Basis für immer wirksamere Therapieverfahren, erfordern aber wiederum immer neue Erklärungen und Theorien größerer Reichweite. Solche immer »wahreren« und immer nützlicheren Theorien wiederum schaffen neue Fakten, die wiederum neue bzw. weiterentwickelte Erklärungen verlangen etc.

■ **Abb. 30.1.** Veranschaulichung des wissenschaftlichen Entwicklungsprozesses der Psychotherapie als Wechselwirkungswirkungsprozess zwischen Theorienentwicklung und Erfahrungsgewinnung. (Grawe, 1995)



Ein Beispiel liefert die systematische Desensibilisierung, deren Wirksamkeit zur Besserung umschriebener Ängste als gesichert angesehen werden kann. Die ursprünglich vom Entwickler dieser Methode (Wolpe, 1958) angenommenen zugrunde liegenden lerntheoretischen Konzepte (u. a. reziproke Hemmung) konnten aber die Wirkung nur unbefriedigend erklären, und der große Erfolg weiterer reizkonfrontationsbasierter Behandlungen (z. B. Flooding) stand dem ursprünglichen Wolpe'schen Erklärungsansatz der notwendigen Gegenkonditionierung sogar diametral entgegen. Schließlich legte Bandura (1977) mit seiner Theorie der Selbstwirksamkeitserwartung eine Theorie zweiter Generation vor, die die bis dahin bekannten verschiedenen Angstbehandlungsmethoden aus einer einheitlichen Perspektive überzeugender erklären konnte als irgendeine der diesen Methoden ursprünglich zugrunde liegenden Theorien. Allerdings ist auch die Reichweite von Banduras Theorie zweiter Generation noch unbefriedigend, da etwa motivationale Aspekte des Erlebens und Verhaltens oder mittlerweile vorliegende biopsychologische Befunde ausgeklammert sind. Doch dies liegt im Grawe'schen Sinne durchaus in der Natur der Sache: Wir stehen angesichts der relativ jungen Geschichte der Interventionsforschung noch am Anfang des Weges von Innovationen über ihre wissenschaftliche Überprüfung hin zu einer neuen Faktenbasis, die wieder neue Entwicklungen hervorbringt, die sich wiederum empirisch bewähren müssen.

Übrigens ist dies in der Psychopharmakologie nicht anders. Auch dort ist man von einer »unifying theory« der Erklärung und Behandlung psychopathologischen Geschehens noch weit entfernt. Man kennt z. B. zwar einige der Mechanismen recht spezifisch, mit denen die Gabe von Serotoninwiederaufnahmehemmern (SSRI) zu einem Rückgang depressiver Symptome führt, aber die Entwicklung neuer Medikamente in diesem Bereich folgt noch immer eher dem Prinzip der kreativen Verbindung von Einzelbefunden. Lassen sich heuristisch daraus Stoffkombinationen erahnen, die z. B. die Nebenwirkungen oder die Latenz (die Zeit bis zum Eintreten der Wirkung) von Antidepressiva reduzieren sollen, werden diese dann anschließend systematisch durchgetestet. Die bis zu der Zulassung eines neuen Medikaments vergleichsweise rigorosen Erfordernisse an Wirksamkeitsnachweise haben zum Ziel, die Weiterentwicklung zumindest hinsichtlich ihrer Wirksamkeit und Unschädlichkeit empirisch abzusichern – auch wenn viele Modelle noch eher korrelativen Charakter haben bzw. eine Theorie der kausalen Erklärung des Gesamtgeschehens bei der Behandlung psychischer Störungen fehlt.

Da sich die Entwicklung klinisch-psychologischer Interventionen wie hier gezeigt in weiten Teilen auch in diesem Stadium befindet, sollen im folgenden Abschnitt

mehrfach methodische Analogien zur Pharmaforschung gezogen werden. Dabei wird besonders der oben angesprochene »konservative« Aspekt betont, d. h. die Absicherung dessen, dass das, was man – auf welcher theoretischen Basis auch immer – entwickelt hat, auch funktioniert (Legitimation und Evaluation). Nach dem daran anschließenden Abschnitt, in dem ein Überblick zum aktuellen Stand der Wirksamkeits- und Effektivitätsforschung (Outcome-Forschung) gegeben wird, werden abschließend die oben genannten »progressiven« Aspekte (Erforschung von prozessualen Wirkmechanismen, gezielte spezifische wissenschaftsinformierte Verbesserungen und Neuentwicklungen) behandelt.

## 30.2 Phasen der Therapieevaluation

### 30.2.1 Vorbemerkung: Wissenschaftliche Fundierung versus wissenschaftliche Anerkennung

Im einleitenden Abschnitt wurde die wissenschaftliche Fundierung von Psychotherapie (und anderer klinisch-psychologischer Interventionen) gekennzeichnet als Wechselspiel zwischen dem innovativen Ausbau einer Therapietheorie (z. B. über ein neues Verständnis gestörter psychischer Prozesse) und dem Nachweis der Wirksamkeit des untersuchten Verfahrens. Buchkremer und Klingberg (2001) stellen in diesem Zusammenhang mehrere Aspekte der Validität eines Therapiekonzepts heraus:

#### Wichtig

Untersuchbar wird der Gegenstand des Interesses erst, wenn die Erreichung expliziter **geeigneter Therapieziele** (d. h. als relevant und als prinzipiell erreichbar erachteten Veränderungen) mit **geeigneten Messverfahren**, die die Störung bzw. deren Veränderung valide und reliabel abbilden, hinsichtlich **definierter Therapierationale** (d. h. theoretisch oder modellhaft abgeleiteter replizierbarer Interventionsstrategien) erforscht wird. Das **Therapieergebnis** (Outcome) sollte dabei möglichst auch hinsichtlich des Therapieprozesses, der der Veränderung zugrunde liegt, beschreibbar sein (Prozess-Outcome-Analysen).

Die Wirksamkeit des Verfahrens ist dabei integraler Bestandteil empirischer Fundierung (auch wenn die genauen Wirkmechanismen noch im Unklaren liegen). Sie muss für jede psychotherapeutische Methode und für verschiedene Störungsbereiche im Einzelnen belegt werden, denn die Aussage »Psychotherapie ist generell wirksam« ist genauso

unwissenschaftlich wie z. B. die Aussage »Chirurgie ist generell wirksam«. Zentrale Anforderungen an solche Studien werden im folgenden Abschnitt zu den Phasen der Psychotherapieprüfung in Analogie zur Pharmaprüfung dargestellt.

Buchkremer und Klingberg (2001) folgend soll jedoch an dieser Stelle darauf hingewiesen werden, dass wissenschaftliche Fundierung als fortlaufender Prozess von der wissenschaftlichen Anerkennung eines Verfahrens (► Kap. 15) zu trennen ist. Letztere ist eine Entweder-oder-Frage im Sinne einer Bestandsaufnahme wissenschaftlicher Evidenz und bedarf einer Konsensbildung innerhalb der wissenschaftlichen Gemeinschaft(en). Zum derzeitigen Konsens gehört, dass Verfahren, die bestimmte Evidenzstufen durchlaufen haben (mehrfacher Wirksamkeitsnachweis in randomisierten kontrollierten Interventionsstudien; »randomized controlled trials«; RCT; ► Abschn. 30.2.3), wissenschaftlich anerkannt werden. Allerdings gibt es auch kritische Gegenstimmen dahingehend, dass der experimentelle Wirksamkeitsnachweis nicht für alle Verfahren die ideale Methodik (Goldstandard) darstellt.

Hierbei sind auch »Zeitgeist« und veränderbare Konventionen sowie gesundheitspolitische Prioritäten und Machtverhältnisse von Bedeutung. Die Rahmenbedingungen wissenschaftlicher und sozialrechtlicher Anerkennung psychotherapeutischer Verfahren (Prüfung durch den »Wissenschaftlichen Beirat Psychotherapie«, WBP, und den »Gemeinsamen Bundesausschuss«, G-BA) finden sich in ► Kap. 15. Die vom WBP eingesetzte Methodik zur Beurteilung der Wissenschaftlichkeit eines Verfahrens sowie alle Gutachten und Diskussionen dazu finden sich unter

<http://www.wbpsychotherapie.de> sowie im Internet-Archiv des Deutschen Ärzteblatts (dort im Archiv als Suchwort WBP eingeben). Neben der seit dem Psychotherapeutengesetz 1999 als Richtlinienverfahren anerkannten Psychoanalyse, der tiefenpsychologisch fundierten Psychotherapie und der Verhaltenstherapie haben mittlerweile auch die Gesprächstherapie und die Systemische Therapie dort wissenschaftliche Anerkennung erfahren – die jedoch wiederum vom G-BA (bislang noch) nicht akzeptiert wurde, weshalb Gesprächs- und Systemische Therapie aktuell keine Regelleistungen der Krankenkassen darstellen.

### 30.2.2 Phasen der Psychotherapieprüfung in Analogie zur Pharmaprüfung

■ Tabelle 30.1 zeigt die vier Phasen der Psychotherapieprüfung in Analogie zur Pharmaprüfung (Müller-Oerlinghausen & Linden, 1981; vgl. Buchkremer & Klingberg, 2001).

In **Phase-I-Studien** werden Psychotherapiekonzeptionen erarbeitet. Dabei wird der Forschungsgegenstand anhand klinischer Erfahrungen, Plausibilität sowie hinsichtlich bereits etablierter störungsspezifischer oder allgemeiner, in Analogstudien ermittelter Befunde beschrieben. Auch Untersuchungen zur Verträglichkeit und Akzeptanz eines Verfahrens sind bereits hier angesiedelt. Unter Zuhilfenahme theoretischer Annahmen und Kasuistiken wird so eine Therapiemethode quasi konstruiert, wobei die Intervention hinsichtlich expliziter Regeln beschreibbar sein muss (z. B. anhand eines Anwendungsmanuals zu Therapieprinzipien und konkreten Interventionen).

■ **Tabelle 30.1.** Phasen der Psychotherapieprüfung in Analogie zur Pharmaprüfung. (Mod. nach Buchkremer & Klingberg, 2001)

Phase	Inhaltlicher Schwerpunkt	Art der Studien
Phase I	Konzeptentwicklung	Explizierung theoretischer Annahmen Analogstudien Kasuistiken Manualentwicklung
Phase II	Exploration mit (nicht kontrollierten) Pilot-Studien	Verlaufsbeschreibende Einzelfallstudien Prä-Post-Analysen in verschiedenen Populationen Offene Studien Durchführbarkeitsstudien Prozess-Outcome-Studien
Phase III	Wirksamkeitsprüfung	Wirksamkeitsstudien (»randomized controlled trials«; RCT) Experimentelle Einzelfallstudien Randomisierte Effektivitätsstudien
Phase IV	Anwendung unter Routinebedingungen, klinische Nützlichkeit	Nicht randomisierte Effektivitätsstudien Erweiterte Einschlusskriterien Zunehmend naturalistische Bedingungen Anwendungsbeobachtungen

In **Phase II** schreitet diese Exploration systematisch fort, indem z. B. bestimmte Patienten mit eng definierten Einschlusskriterien in Einzelfall- bzw. Zeitreihenanalysen beobachtet werden. Die Methodik ist hier noch nicht streng experimentell kontrolliert und folgt den im Einzelfall angemessenen Erfordernissen für Plausibilitätsprüfungen und der Exploration von möglichen Risiken und Nebenwirkungen.

In **Phase III** kommt die nun mittlerweile gut elaborierte Intervention richtig »auf den Prüfstand«: Die Wirksamkeit muss hier konfirmativ belegt werden (erweiterte klinische

Prüfung). Da diese Prüfung für den wissenschaftlichen Stellenwert eines Verfahrens von besonderer Bedeutung ist, wird ihr unten ein eigener Abschnitt gewidmet. Studien der Phase III sind in der Regel prospektive, randomisierte, kontrollierte Interventionsstudien (»randomized controlled trials«, RCT) an einer klar definierten Zielgruppe mit hohen methodischen Anforderungen (z. B. Poweranalyse, Drop-out-Analyse, Kontrolle von Selektionseffekten, möglichst zumindest mittelfristige Katamnesen). Begleitend sollte dabei nach Möglichkeit zur Erfassung potenzieller Wirkfaktoren zusätzlich eine Prozess-Outcome-Forschung

### Gut zu wissen

#### Das Spannungsfeld »interne versus externe Validität«

Martin Seligman stellt in seinem Artikel »The Effectiveness of Psychotherapy« (1995) die Unterschiede von Wirksamkeits- und Effektivitätsstudien heraus und macht Vorschläge, wie diese beiden Forschungsperspektiven ergänzend eingesetzt werden könnten (Phase III und Phase IV). Dazu gehört es, in naturalistischen Versorgungsstudien, wie sie diese z. B. die amerikanische Verbraucherorganisation Consumer Reports (vergleichbar der deutschen Stiftung Warentest) in einer großen Umfrage zur Inanspruchnahme und zu den Effekten psychosozialer Dienste durchgeführt hat, auf einige der strengen Kriterien experimenteller und quasi-experimenteller Studien zur Sicherstellung hoher interner Validität zu verzichten. Denn die externe bzw. ökologische Validität könnte z. B. durch die Auslese der Patienten ohne Komorbidität, Manualisierung, standardisierte Therapiedauer, Kontrollgruppen mit Zufallszuweisung etc. gefährdet werden. Außerdem sollten die Ergebnisse aus Therapiestudien von Forschergruppen repliziert werden, die nicht an der Entwicklung der zu überprüfenden Intervention beteiligt waren, denn besonders starke Identifikation mit und eine außergewöhnliche Expertise in dem beforschten Ansatz können zu unrealistisch hohen Effekten führen. Dieses Problem wird allgemein als »Allegiance-Effekt« bezeichnet (»allegiance« = Gefolgschaft, Hörigkeit, Untertanentreue). Auch Shadish, Matt, Navarro und Siegle (1997) problematisieren, dass Psychotherapiestudien nur selten unter klinisch repräsentativen Bedingungen stattfinden, stellen aber in einer späteren Metaanalyse (► Abschn. 30.3.1) fest, dass die Effekte unter klinisch repräsentativen Bedingungen ähnlich denen aus kontrollierten Wirksamkeitsstudien sind (Shadish, Matt, Navarro & Philips, 2000). Allgemein werden Forderungen laut, es fehle an Studien aus dem Feld, in denen spezifische, für den Therapieerfolg möglicherweise relevante Randbedingungen im Hinblick auf ihre Kausalwirkung für den Therapieerfolg untersucht werden können (Höfler, Gloster & Hoyer, 2010).

Doch Vorsicht – hierbei dürfen nicht die Grenzen der Interpretierbarkeit übersehen werden, die naturalistischen Studien wie bei Consumer Reports, die sich ausschließlich auf eine hohe externe Validität konzentrieren, oft zu eigen sind: Vor allem die Selbstselektion der Stichprobe und die subjektive Post-hoc-Befragung als einzige Datenquelle stellen nämlich wiederum eine drastische Bedrohung der internen Validität dar. Eine derartige Studie hätte heutzutage kaum eine Chance, im Rahmen der wissenschaftlichen Anerkennung eines Verfahrens oder der Aufnahme eines Verfahrens in eine Liste von »empirically supported treatments« (► Abschn. 30.3.2) als Wirksamkeitsbeleg herangezogen zu werden.

Weiterhin sind im Zusammenhang mit den verschiedenen Validitätskriterien auch statistische Fragen von Bedeutung. Therapiestudien mit unausgelesenen Patienten im naturalistischen Setting sind meist mit einer hohen Heterogenität der untersuchten Patienten und Behandlungen verbunden (z. B. aufgrund Komorbidität oder unterschiedlicher Therapiedauer) und damit auch mit einer hohen Varianz der Ausgangsdaten. Damit sinkt jedoch die Wahrscheinlichkeit, statistisch signifikante Ergebnisse zu finden. Es können aber weder die Effekte einer Intervention noch die Stichprobe beliebig vergrößert werden. Eine versuchsplanerische Homogenisierung zur Erhöhung der Teststärke ist also bisweilen notwendig, um eine realistische Chance zu behalten, reell vorhandene Effekte aufzudecken. Allerdings gibt es immer wieder Effectiveness-Studien mit großer Stichprobenheterogenität, bei denen die Effektstärken, die in Efficacy-Studien gefunden wurden, nicht abfallen (z. B. Lincoln et al. 2003; Shadish et al. 2000).

Die unterschiedlichen Herangehensweisen im Spannungsfeld von interner und externer Validität hängen somit von der jeweiligen Fragestellung ab und sollten sich komplementär ergänzen – im Zweifelsfall ist allerdings ein notwendiges Minimum an interner Validität einem maximalen Anspruch auf externe Validität vorzuziehen.

durchgeführt werden (vgl. Borkovec & Castonguay, 1998). Prinzipiell können auch experimentelle Einzelfallstudien in dieser Phase als Wirksamkeitsnachweis dienen; Publikationen dieses Forschungsansatzes sind allerdings bislang in Fachzeitschriften eher unterrepräsentiert (Perst & Baumann, 1999). Dass aber Einzelfallstudien im Rahmen der Entwicklung und Beurteilung von Interventionsverfahren eine wichtige Bedeutung haben, wird u. a. dadurch dokumentiert, dass entsprechende Publikationen zunehmen. Mittlerweile gibt es spezielle Fachzeitschriften für diesen Bereich, z. B. »Clinical Case Studies« (seit 2002).

Nachdem eine Reihe von Wirksamkeitsnachweisen (über verschiedene Studienzentren und Störungsbereiche hinweg) aus der Phase III vorliegen, wird ein Verfahren in der Regel anerkannt bzw. ein neues Medikament zugelassen. Das Bewertungsverfahren der berufrechtlichen und sozialrechtlichen Anerkennung nach Prinzipien der »evidence based medicine« (Sackett, Rosenberg, Gray, Haynes & Richardson, 1996) in Bezug auf psychologische Behandlungen in Deutschland ist in ► Kap. 15 zur Versorgung psychischer Störungen beschrieben.

In der folgenden **Phase IV** wird die klinische Nützlichkeit bzw. Effektivität bei meist erheblich vergrößerter Fallzahl unter Praxisbedingungen (»effectiveness«) untersucht, wobei die strengen Kontrollen der Wirksamkeitsstudien (»efficacy«; z. B. enge Einschlusskriterien, Zufallszuweisung, standardisierte Therapiedauer bzw. -dosis) systematisch gelockert werden. Neben dem Nachweis von Verbesserungen hinsichtlich einer umgrenzten Zielsymptomatik werden hier auch breitere, sekundäre Konstrukte berücksichtigt, wie etwa allgemeine Lebensqualität oder Kosten-Nutzen-Aspekte. Somit wird die externe bzw. ökologische Validität erhöht, die noch in Phase III hinter die Maximierung der internen Validität zurücktreten musste. Obwohl sich diese beiden Validitätsaspekte komplementär ergänzen, ist es im Sinne des »Verbraucherschutzes« sinnvoll, ein neues Verfahren erst dann zu etablieren, wenn es die Phase III erfolgreich durchlaufen hat – ein Überspringen dieser strengen Prüfung zugunsten einer von vornherein »naturalistisch« angelegten Forschungsstrategie entspricht nicht der derzeitigen (und wohl auch noch auf absehbare Zeit üblichen) Anerkennungspraxis.

### 30.2.3 Die besondere Bedeutung der Phase III: Anforderungen an Wirksamkeitsstudien

Die Kriterien für den Prozess der Entwicklung und Beurteilung von Pharmaka im Sinne einer »good clinical practice« und die methodisch-statistischen Prinzipien

für entsprechende Phase-III-Untersuchungen sind als Richtlinien mit internationalem Konsens differenziert dargelegt (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Humans, 1998). Die umfangreiche Vorarbeit, die in diesem Ansatz steckt, kann sich die klinisch-psychologische Interventionsforschung nutzbar machen, denn bis auf Ausnahmen sind die Prinzipien für die Sicherstellung hoher Studienqualität gut auf unseren Bereich übertragbar; sie entsprechen letztendlich übergreifenden wissenschaftlichen Standards (insofern man sich im Forschungsfeld von Wirksamkeitsstudien bewegt, denn nicht alle relevanten Fragestellungen sind auf diese Weise untersuchbar; ► Abschn. 30.3). Da einer der wichtigsten dieser Standards darin besteht, dass die Studienergebnisse vor dem Hintergrund von Studiendesign und Studiendurchführung transparent und intersubjektiv nachvollziehbar berichtet werden müssen, existiert außerdem ein verbindlicher Leitfaden gemeinsamer Standards für die Beschreibung von Studien (Consolidated Standards of Reporting Trials, CONSORT Statement; dt. Moher, Schulz, Altman für die CONSORT-Gruppe, 2004). Da empirische Evidenz gezeigt hat, dass das Nichtberichten von Information mit verzerrten Schätzungen des Behandlungseffektes verbunden ist, findet sich dort eine Checkliste mit 22 Punkten, über die die Studienautoren verbindlich berichten müssen (■ Tab. 30.2).

Besonders wichtig ist hier etwa die genaue Dokumentation aller einbezogenen Fälle (einschließlich Abbrecher und nicht auswertbare Fälle zu verschiedenen Zeitpunkten; ■ Abb. 30.2).

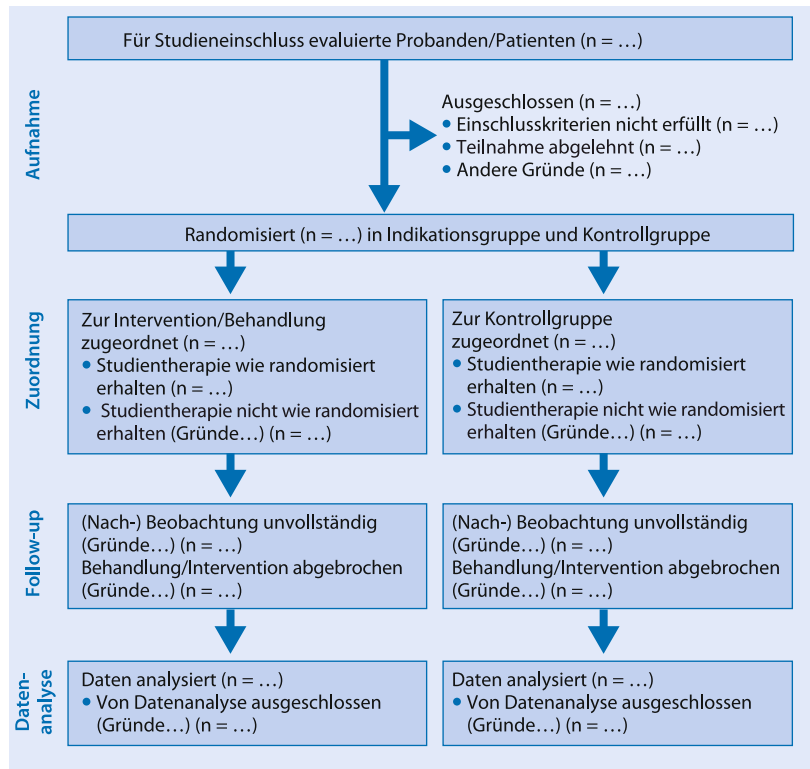
Im Folgenden sind entlang des Ablaufs einer Phase-III-Studie die wichtigsten Prinzipien und Begriffe dieser wissenschaftlichen Leitlinien aufgeführt, ergänzt um Besonderheiten für den Bereich klinisch-psychologischer Studien. Die Lektüre solcher Methoden-Leitfäden (wie die im Folgenden zusammengefasst: Guideline E9: Statistical Principles for Clinical Trials; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Humans, 1998) ist nützlich bzw. notwendig, wenn man an einer Interventionsstudie beteiligt ist (auch an einer außerhalb der Phase III). Sie kann aber auch generell zu Übungszwecken ausdrücklich empfohlen werden. Da auch in diesem Bereich die Internationalität voranschreitet – selbst Forschungsanträge für deutsche Therapiestudien müssen auf Englisch verfasst werden –, werden im Folgenden die englischen Termini immer mit erwähnt.



■ **Tabelle 30.2.** Checkliste zur Publikation randomisierter Studien. (Nach Moher et al. 2004)

Publikationsabschnitt	Beschreibung
<b>Titel und Zusammenfassung</b>	1 Zuordnung zu Therapiegruppen (z. B. »randomisierte Verteilung«, »randomisiert« oder »randomisierte Zuweisung«)
<b>Einleitung</b>	
Hintergrund	2 Wissenschaftlicher Hintergrund und Begründung der Studie
<b>Methoden</b>	
Probanden/Patienten	3 Einschlusskriterien der Probanden/Patienten; Studienorganisation und Ort der Studiendurchführung (z. B. im Krankenhaus oder nicht-stationär)
Intervention/Behandlung	4 Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Durchführung
Ziele	5 Genaue Ziele, Fragestellungen und Hypothesen
Zielkriterien	6 Klar definierte primäre und sekundäre Zielkriterien und, gegebenenfalls, alle zur Optimierung der Ergebnisqualität verwendeten Methoden (z. B. Mehrfachbeobachtungen, Training der Prüfer)
Fallzahlbestimmung	7 Wie wurden die Fallzahlen bestimmt und, falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch
Randomisierung	
Erzeugung der Behandlungsfolge	8 Methode zur Generierung der zufälligen Zuteilung, einschließlich aller Einzelheiten (wie z. B. Blockrandomisierung, Stratifizierung)
Geheimhaltung der Behandlungsfolge (»allocation concealment«)	9 Durchführung der Zuteilung (z. B. nummerierte Behälter, zentrale Randomisierung per Fax/Telefon); Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war
Durchführung	10 Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?
Verblindung	11 Waren a) die Probanden/Patienten und/oder b) diejenigen, die die Intervention/Behandlung durchführten und/oder c) diejenigen, die die Zielgrößen beurteilten verblindet oder nicht verblindet? Wie wurde der Erfolg der Verblindung evaluiert?
Statistische Methoden	12 Statistische Methoden zur Bewertung des primären Zielkriteriums; weitere Analysen wie z. B. Subgruppenanalysen und adjustierte Analysen
<b>Ergebnisse</b>	
Ein- und Ausschlüsse	13 Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) tatsächlich die geplante Behandlung/Intervention erhalten haben, b) die Studie protokollgemäß beendeten, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden (Darstellung in Flussdiagramm empfohlen; Beschreibung von Protokollabweichungen mit Angabe von Gründen)
Aufnahme/Rekrutierung	14 Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden/Patienten und der Nachbeobachtung
Patientencharakteristika zu Studienbeginn (»baseline data«)	15 Demographische und klinische Charakteristika aller Gruppen
Anzahl der ausgewerteten Probanden/Patienten	16 Anzahl der Probanden/Patienten (Nenner) in jeder Gruppe, die in die entsprechende Analyse eingeschlossen wurden, und Angabe, ob es sich dabei um eine »Intention-to-treat«-Analyse handelt. Wenn möglich, Angabe der Ergebnisse in absoluten Zahlen (z. B. 10 von 20, nicht 50%)
Ergebnisse und Schätzmethoden	17 Zusammenfassung der Ergebnisse aller primären und sekundären Zielkriterien für jede Gruppe und die geschätzte Effektgröße sowie ihre Präzision (z. B. 95%-Konfidenzintervall)
Zusätzliche Analysen	18 Angabe von weiteren Tests, insbesondere von Subgruppenanalysen und adjustierten Analysen (mit Erklärung, ob sie vorher geplant waren oder nachträglich durchgeführt wurden)
Unerwünschte Wirkungen	19 Angabe der wichtigen unerwünschten Wirkungen oder Nebenwirkungen innerhalb jeder Behandlungsgruppe
<b>Diskussion</b>	
Interpretation	20 Interpretation der Ergebnisse unter Berücksichtigung der Studienhypothesen, möglicher Ursachen von Verzerrungen (»bias«) sowie Problemen durch multiples Testen und multiple Zielkriterien
Generalisierbarkeit	21 Generalisierbarkeit der Studienergebnisse (externe Validität)
Bewertung der Evidenz	22 Allgemeine Interpretation der Ergebnisse unter Berücksichtigung des aktuellen Forschungsstandes und anderer Publikationen zur untersuchten Fragestellung

**Abb. 30.2.** CONSORT Statement: Flussdiagramm zur Stichprobenbeschreibung und Verteilung der Ausfälle im Verlauf einer randomisierten Studie. (Nach Moher et al. 2004)



### Vorab: Was wird denn hier eigentlich hinsichtlich der Wirksamkeit getestet?

Wenn eine Studie als »Vergleich zwischen kognitiver Verhaltenstherapie (KBT, engl. CBT) und der medikamentösen Therapie mit Serotoninwiederaufnahmehemmern (SSRI) bei Depression« deklariert wird, scheint man zunächst sofort zu wissen, worum es geht. Doch beide Arten der Intervention können unter derselben Überschrift höchst unterschiedlich ausgestaltet sein, z. B.: Welcher SSRI genau? In welcher Dosis – bei allen Patienten die gleiche? Bekommt der Patient einfach die SSRI in die Hand oder gibt es darüber hinaus noch Psychoedukation, schriftliches Material, klinisches Management, Beratung und Ermunterung durch die Studienärzte – und wenn ja, in welchem Umfang und wie standardisiert? Gleiches gilt für das Label »CBT«: Auch hier muss – in der Regel in Form umfangreicher Manuale – genau festgelegt und transparent gemacht werden, wie die Behandlung genau vonstatten geht und wie sichergestellt wird, dass die Studientherapeuten sich auch an das Manual halten. Ferner muss genau geprüft werden, ob bzw. in welchem Umfang sich die Patienten auch an die therapeutischen Vorgaben gehalten haben. Zu verstehen und angemessen darzustellen, was eigentlich in den Studientherapien genau passiert ist, ist also alles andere als trivial.

Die gesamte Umsetzung der intendierten Behandlungsmaßnahme (»treatment fidelity«) sowohl auf Therapeuten-seite (»Manualtreue«, »treatment integrity«, »treatment delivery« oder »adherence«) als auch auf Patientenseite (»treatment receipt«, »enactment«, »compliance«) zu beschreiben, zu bewerten und schließlich auch die gefundenen Therapieergebnisse vor dem Hintergrund der realisierten Treatment-Fidelity angemessen zu interpretieren, ist also im Grunde eine Studie für sich. Videoanalysen, penibel geführte Dokumentation jeder einzelnen Studientherapie und fortlaufende Supervision stellen höchste Anforderungen der Qualitätssicherung dar – auch wenn am Ende in der zentralen Publikation der Therapiestudie dazu nur wenige aussagekräftige Sätze im Methodenteil stehen.

Beispielhafte Maßnahmen zur Sicherstellung bzw. Kontrolle der Treatment-Fidelity, wie sie etwa in einer großen, durch das BMBF geförderten Multicenterstudie zur Behandlung der Panikstörung (s. <http://www.paniknetz.de>) getroffen wurden, sind:

- ausführliche Manualisierung jeder Sitzung und des gesamten Therapieablaufs mit genauer Operationalisierung von Gütekriterien kompetenter Umsetzung, so dass im nächsten Schritt diese Umsetzung anhand von Videoaufzeichnungen systematisch bewertet werden kann;

- umfangreiche Schulung und Zertifizierung der Studientherapeuten, sowie Auffrischung/Nachschulungen, um die Kontinuität der Adherence aufrechtzuerhalten;
- kontinuierliche Supervision (mit verpflichtender Video- und Dokumentationsanalyse);
- umfangreiche Dokumentation jeder einzelnen Sitzung (einschließlich Bewertung der Manualtreue, Umsetzung von Hausaufgaben, Einschätzung des Verständnisses der Inhalte auf Patientenseite etc.);
- und nicht zuletzt: Das durch diese qualitätssichernden Maßnahmen entstehende und äußerst umfangreiche Datenmaterial muss auch zeitnah ausgewertet werden!

Eine differenzierte Anleitung und Empfehlungen zur »best practice« hierzu liefert das NIH Behavior Change Consortium (Bellg et al. 2004).

### Allgemeine Prinzipien für Wirksamkeitsstudien (auch: Bestätigungstest)

Wichtig		
■	Untersuchungsziel: Wirksamkeit und Verträglichkeit	
■	Transparenz und Nachvollziehbarkeit sicherstellen	
■	Prospektiv angelegtes Studienhandbuch (»protocol«)	
■	Bias minimieren, Präzision maximieren	
■	Abschätzen, wie robust Ergebnisse sind	
■	Begutachtung durch Ethikkommission (zentral: vorläufige Unbedenklichkeit, »informed consent«, Datenschutz)	

Um das Studienziel (Erkenntnisfortschritt hinsichtlich Wirksamkeit und Verträglichkeit einer oder mehrerer Interventionen) zu erreichen, müssen die Ergebnisse auf einer sauberen Methodik basieren (z. B. präzise und ohne Bias sein) und entsprechend nachvollziehbar dokumentiert und berichtet werden (z. B. Angaben dazu machen, für wie robust die Ergebnisse erachtet werden). Zentrales Element ist das Studienhandbuch (»protocol«), in dem *vorab* alle Fragestellungen, Hypothesen, Designaspekte, Messmethoden, Auswertungsstrategien, Erfolgskriterien festgelegt werden. Neben dem Studienleiter (»principal investigator«, PI) übernimmt hierbei ein möglichst eigens zu bestellender Methodiker/Statistiker besondere Verantwortung. *Jegliche* Änderungen im geplanten Ablauf (z. B. Änderung der Ein- und Ausschlusskriterien oder neue Rekrutierungstechniken, wenn sich im Rahmen der Behandlung herausstellt, dass sich mit der bisherigen Konzeption die angestrebte Patientenzahl nicht erreichen lässt) sind fortlaufend in diesem Studienhandbuch zu dokumentieren.

Vor Studienbeginn ist eine Stellungnahme der zuständigen Ethikkommission einzuholen. Gegebenenfalls müssen daraufhin noch entsprechende Änderungen im Studienablauf vorgenommen werden. Wichtig ist in diesem Zusammenhang, zu zeigen, dass bei der untersuchten Intervention mit einem akzeptablen Risiko-Nutzen-Verhältnis zu rechnen ist. Die Probanden müssen detailliert über die Studienbedingungen (z. B. mögliche Zuweisung zu einer Placebobedingung oder zu einer Gruppe, die eine noch ungeprüfte Intervention erhält) und mögliche Risiken informiert werden und ihr schriftliches Einverständnis zur Studienteilnahme geben (»informed consent«). Ausführlich ist auch darzulegen, wie alle datenschutzrelevanten Standards erfüllt werden. Zum Beispiel müsste in einer Studie zur Behandlung von Substanzstörungen, bei der die Patienten u. a. auch nach Drogenkonsum gescreent werden, in der Patienteninformation detailliert dargestellt werden, wer dieses Screening durchführt, auf welche Drogen getestet wird, wie die Urinproben gekennzeichnet werden, was damit am Ende der Untersuchung passiert, in wie weit für die Personen, die Umgang mit diesen sensiblen personenbezogenen (ggf. strafrechtlich relevanten) Daten eine Schweigepflicht bzw. ein Zeugnisverweigerungsrecht gilt, wo die Daten gespeichert werden, was mit den Daten nach Studienabschluss passiert (vollständige Anonymisierung?) etc.

### Inhaltliche Festlegung der Outcome-Variablen

Wichtig		
■	Primäre Variablen (»target variable«, »primary endpoint«)	
■	Sekundäre Variablen	

Eine primäre oder Zielvariable sollte so gewählt werden, dass das primäre Studienziel theoretisch und klinisch-praktisch bedeutsam beurteilt werden kann (z. B. bei einer Studie im Suchtbereich: Rate der nach dem festgesetzten Untersuchungszeitraum abstinenten Behandelten; bei Depression: Rate derer, deren Wert auf Depressionsskala XY sich zumindest halbiert hat; vgl. Hiller, Bleichhardt & Schindler, 2009, sowie den Exkurs zu Effektstärken und klinischer Signifikanz in ► Abschn. 30.2.4). Während im Pharmabereich (sekundäre) Variablen zu beschränken, besteht im Rahmen der Psychotherapieforschung traditionell der Konsens, dass die Effekte möglichst breit und multimethodal erfasst werden sollen – hierbei sind aber mögliche methodische bzw. statistische Probleme (z. B. Abhängigkeit verschiedener Maße, »multiplicity problems«) vorher zu berücksichtigen und im Studienhandbuch zu antizipieren.

## Form der Outcome-Variablen

Wichtig		
■	Kategorial versus kontinuierlich	
■	Einzel/eindimensional	
■	Zusammengesetzt (»composite«)	
■	Global	

Neben der eben genannten dichotomen Betrachtung (z. B. Prozentanteil geheilter Patienten) können auch dimensionale Maße (z. B. mittlere Reduktion der Werte auf einer Angstskala) eingesetzt werden. Möchte man mehrere (voneinander nicht 100% unabhängige) Bereiche inhaltlich abdecken, ohne mit der Abhängigkeit der Werte zu kämpfen, bieten sich zusammengesetzte Variablen an, die hinsichtlich eines zuvor festgelegten Algorithmus verrechnet werden. Globale Variablen, bei denen Patienten und Therapeuten (oder externe Beurteiler) im Sinne einer direkten Veränderungsmessung eine Gesamteinschätzung abgeben, haben den Vorteil, dass eine breite Beurteilung vorgenommen wird (in die z. B. auch Nebenwirkungen einfließen). Sie haben aber oft schlechte psychometrische Eigenschaften (z. B. wenig Varianz, nicht normalverteilt) und sind als primäre Variablen ungeeignet, da sich die verschiedenen Einflüsse, die in sie eingehen, später nicht mehr auseinanderdividieren lassen.

## Kontrollgruppen

Wichtig		
■	Placebo	
■	Keine Behandlung	
■	Unterschiedliche Behandlungsintensität (Dosis)	
■	Andere aktive Behandlung	

Es gibt verschiedene Kontrollbedingungen, anhand derer abgeschätzt werden kann, welche Veränderungen auch ohne die im Fokus stehende Intervention stattgefunden hätten. Eine glaubwürdige (»credible«) klinisch-psychologische Placebobedingung (z. B. »unspezifische stützende Gespräche ohne therapeutische Intervention«) ist schwerer zu realisieren als im pharmakologischen Bereich. Eine Übersicht zum derzeitigen Stand der Erklärungsmöglichkeiten für Placeboeffekte findet sich bei Stewart-Williams (2004). Placeboeffekte sollten übrigens in jedem Fall ernst genommen und bei der Bewertung der Ergebnisse berücksichtigt werden (auch wenn in der eigenen Studie keine Placebobedingung realisiert wurde), denn ein erheblicher Teil der gefundenen Effekte in der »aktiven Interventionsgruppe« können auf Placeboeffekte zurückgeführt werden;

Rief et al. (2009) stellten bei der Analyse von Placeboeffekten in Wirksamkeitsstudien zu Antidepressiva fest, dass dieser Anteil 68% beträgt! Demnach sollten Placeboeffekte in klinischen Studien also unbedingt kontrolliert, aber wiederum in der klinischen Praxis voll ausgeschöpft bzw. genutzt werden (► Abschn. 30.4.1).

»Keine Behandlung« wird oft als Wartegruppenbedingung realisiert (d. h. den Teilnehmern wird zwar eine Behandlung angeboten, aber verzögert; Kontrollbedingung ist die Zeit vor der Intervention), wobei im Einzelfall abgeschätzt werden muss, in wie weit auch die Aufnahme in eine Warteliste bereits eine Intervention mit möglichen Effekten darstellt.

Dosis-Wirkungs-Beziehungen sind nicht nur in der Pharmakologie, sondern auch in Bezug auf klinisch-psychologische Interventionen interessant (z. B. eine »minimale Intervention« mit 5 Stunden und umfangreichem Selbsthilfematerial versus »umfangreiche Intervention« mit 25 Stunden zu den gleichen Inhalten).

Wird die Überlegenheit gegenüber einer bereits als wirksam erachteten Behandlung nachgewiesen, ist dies ein besonders überzeugender Wirksamkeitsbeleg. Zudem kann bei einem solchen Design die andere aktive Bedingung systematisch so ausgewählt werden, dass sich auch Analysen dazu anstellen lassen, was (bei einem Bündel von Einzelinterventionen und Therapiebausteinen) besonders zur Gesamtwirksamkeit beigetragen hat.

## Minimierung von systematischen Verzerrungen

Wichtig		
■	Blind, doppelblind	
■	Randomisierung	

Die wichtigsten Techniken zur Vermeidung von systematischen Verzerrungen sind Randomisierung und Verblindung (d. h. Patient weiß nicht, welches Treatment er erhält; doppelblind: auch der Untersucher ist im Unklaren). Bei Psychotherapiestudien ist dies oft nur begrenzt möglich. Zum Beispiel weiß der behandelnde Therapeut immer, welche Behandlung er einsetzt, im Gegensatz zu einer Pharmastudie, in der aktive und Placebopillen äußerlich nicht zu unterscheiden sind. Allerdings sind auch in Pharmastudien Teilnehmer und Forscher oft nicht wirklich blind, da sie Hypothesen aufgrund der erlebten (Nicht-)Wirkung anstellen (vgl. Margraf et al. 1991). Optimal ist in diesem Bereich, zumindest die Ergebnisbeurteilung von externen Evaluatoren vornehmen zu lassen, denen der Behandlungsstatus des Patienten unbekannt ist.

Auch die Methode der Randomisierung muss bestimmten Regeln folgen und vorher genau spezifiziert wer-

den (optimal: Randomisierung vom technischen Personal am Telefon; ungenügend: bei Anmeldung persönliche Zuordnung durch den Therapeuten zu der Gruppe, in der »zufällig« gerade Platz ist). Insbesondere in kleineren Studien ist es u. U. angebracht, blockweise zu randomisieren, um die Untersuchungsgruppen möglichst vergleichbar zu halten. Bei Vorliegen mehrerer Studienzentren sollte die Randomisierung zentral und möglichst hinsichtlich prognostisch wichtiger Faktoren stratifiziert erfolgen (d. h. Balancierung der Stichprobe über die verschiedenen Zentren hinweg).

### Designtypen

Wichtig		
■	Parallelgruppendesign	
■	Crossover-Design	
■	Faktorielles Design	
■	Multicenterstudie	

Das einfachste Design, bei dem die Teilnehmer einfach zufällig zur jeweiligen Gruppe zugeordnet werden, ist das Parallelgruppendesign. Crossover-Designs (z. B. Wechsel zwischen aktiver und Placebobehandlung) sind bei klinisch-psychologischen Interventionsstudien in der Regel nicht angemessen, da hier mehr noch als im pharmakologischen Bereich »Carry-over-Effekte« zu erwarten sind (d. h. eine psychologische Intervention kann meist nicht so einfach »ausgeschlichen« werden wie ein Medikament).

Beim faktoriellen Design werden mehrere aktive Treatments (und ggf. deren Kombination) in einer Studie hinsichtlich ihrer Wirksamkeit untersucht (► Kasten »Was wirkt besser: Medikamente oder Verhaltenstherapie?«, S. 626).

Multicenterstudien bergen neben dem praktischen Vorteil, dass mehrere Untersuchungsstandorte mehr Probanden rekrutieren können, auch größere Generalisierungsmöglichkeiten in sich; zudem kann insbesondere bei therapeutisch unterschiedlich ausgerichteten Forschergruppen der sog. Allegiance-Effekt (Bias in Richtung eigener therapeutischer Präferenz) minimiert werden. Selbstverständlich muss hier in besonderer Weise sichergestellt werden, dass die spezifizierten Behandlungen über die Studienzentren hinweg korrekt durchgeführt und strikt vergleichbar sind (»treatment integrity«). Wie mit Heterogenität der Ergebnisse über die Zentren hinweg umgegangen wird (z. B. getrennte Auswertung, Einführung von Interaktionstermen: Treatment × Zentrum), muss vorab festgelegt werden.

### Vergleichsmöglichkeiten

Wichtig		
■	Nachweis der Überlegenheit (»superiority trial«)	
■	Nachweis der Äquivalenz oder Nichtunterlegenheit (»non-inferiority trial«)	
■	Nachweis einer Dosis-Wirkungs-Beziehung	

Der überzeugendste Nachweis der Wirksamkeit ist der Nachweis der Überlegenheit gegenüber einer Placebo- oder aktiven Kontrollbedingung. In manchen Fällen (z. B. im Bereich der Neuroleptika für Schizophrenie) ist eine Placebobedingung ethisch nicht vertretbar (und würde auch die Verblindung aufheben, denn die Nichtwirkung eines Placebos kann hier offensichtlich sein). Bei klinisch-psychologischen Interventionsstudien bietet sich der Vergleich mehrerer aktiver Bedingungen auch deswegen an, weil hier Placebobedingungen meist grundsätzlich schwerer herzustellen und schwerer beurteilen zu sind. Außerdem kann hier zusätzlich zum Wirksamkeitsnachweis systematische Prozess-Outcome-Forschung betrieben werden (z. B. Hinzufügen und Entfernen von Therapiebausteinen).

Eine denkbare Variante eines Nachweises der Äquivalenz oder Nichtunterlegenheit (eher im pharmakologischen Bereich anzutreffen) wäre eine Kosten-Effektivitäts-Studie, bei der eine herkömmliche und eine minimale (d. h. deutlich billigere) Intervention miteinander konkurrieren, da u. U. gar nicht beabsichtigt ist, die Effektivität zu steigern, sondern eher, die Kosten bei gleichem Resultat zu senken. Selbstverständlich müssen hierbei in besonderem Ausmaß methodische Risiken beachtet werden, denn bereits statistisch ist ein Test der Nichtunterlegenheit vom Wesen her nicht konservativ. Außerdem wäre in unserem Beispiel genau zu beachten, ob nicht vielleicht die billigere Variante schlechte Langzeiteffekte aufweist. Gerade in diesem heiklen Bereich ist die prospektive Dokumentation im Studienhandbuch durch den Statistiker obligatorisch.

Auch der Nachweis einer Dosis-Wirkungs-Beziehung gilt als Wirksamkeitsbeweis; außerdem ermöglicht ein solcher Ansatz die Entwicklung von Dosis-Wirkungs-Profilen, die als Referenz für konkrete Einzelverläufe dienen können (► Abschn. 30.2.4).

### Bedeutung der Stichprobengröße

Die grundlegenden Kenntnisse zum Zusammenhang zwischen  $\alpha$ -Niveau, Teststärke und Stichprobengröße werden hier als bekannt vorausgesetzt. In klinischen Studien gilt es aber zu beachten, dass bei bestimmten Konstellationen, z. B. bei einer Vielzahl primärer und sekundärer Auswer-

tungen, oder wenn der Fokus auf Unbedenklichkeit und Verträglichkeit liegt (unerwünschte Ereignisse, »adverse events«, sind meist selten), größere Stichproben benötigt werden. Weitere Besonderheiten in klinischen Studien ergeben sich hinsichtlich der Stichprobengrößen bei Dosis-Wirkungs-Zusammenhängen. Wie in der Beispielstudie im ► Kasten »Was wirkt besser: Medikamente oder Verhaltenstherapie?« dargestellt, können auch differenzielle Drop-out-Raten sowohl die Berechnung als auch die Interpretation der Ergebnisse verkomplizieren.

Bevor die Rekrutierung für die Studie startet, sollten sämtliche für das Thema Stichprobengröße relevanten Modellrechnungen und Kalkulationen im Studienhandbuch stehen.

### Steuerung des Studienablaufs (»trial monitoring«) und Zwischenauswertungen (Interimsanalysen)

Die fortlaufende Überwachung des Studienablaufs soll gewährleisten, dass das Protokoll eingehalten und die Datensammlung optimiert wird. Müssen vorab im Studienhandbuch festgelegte Abläufe verändert werden, muss dies mit Begründung dokumentiert werden (z. B. Änderung der Ein- und Ausschlusskriterien, weil sich während der Rekrutierung Schwierigkeiten ergeben haben oder Erhöhung der Stichprobengröße, weil über alle Untersuchungsbedingungen hinweg die Effekte kleiner ausfallen als bei der ursprünglichen Poweranalyse angenommen).

Allgemeine Qualitäts-Checks wie die Minimierung fehlender Werte (»missings«) und die Verhinderung von Ausfällen (»drop-outs«) bei den Messungen, oder Plausibilitätsprüfung der Daten sind dabei von Zwischenauswertungen hinsichtlich der Studienziele abzugrenzen. Solche Interimsanalysen dienen dazu, mögliche unerwünschte Effekte der Intervention frühzeitig zu erkennen, oder ggf. die Studie vorzeitig abubrechen, wenn absehbar ist, dass die untersuchte Intervention auch bei Erreichen der endgültigen Stichprobengröße keine Chance mehr hätte, ihre Überlegenheit zu zeigen. Zwischenauswertungen werden in der Forschungspraxis häufig vorgenommen (und dann z. B. auf Kongressen als Poster präsentiert), obwohl dies den Forschungsleitlinien widerspricht – eigentlich sollten sie sparsam eingesetzt und sorgfältig mit genauer Rollenverteilung (z. B. unabhängige Auswerter) geplant werden. Denn wenn Versuchsleiter und Behandler erste Tendenzen der Ergebnisse vor Abschluss der Studie mitgeteilt bekommen, könnte dies u. U. deren Einstellung und Verhalten im Versuch – bewusst oder unbewusst – beeinflussen und damit die Ergebnisse verfälschen (z. B. Veränderungen bei Rekrutierung und Behandlungsdurchführung).

## Datenanalyse

### Wichtig

- Gesamtgruppe (»full analysis set«) und »Intention-to-treat«-Prinzip
- Vollständige Datensätze (»per protocol set«, »valid cases«)
- Spezifikation des Hypothesentests

Wenn alle Teilnehmer einer Wirksamkeitsstudie die Ein- und Ausschlusskriterien bis zum Studienabschluss erfüllen, alle Abläufe in ihrer jeweiligen Untersuchungsbedingung wie geplant durchlaufen, und wenn es keine Abbrecher gibt und alle Daten bis zum Follow-up vollständig sind, dann wäre die Analysestichprobe eindeutig festgelegt. In der Praxis sind aber solche idealen Bedingungen in der Regel nicht gegeben, sodass bei der Analyse sehr sorgfältig und transparent vorgegangen werden muss. ► Abb. 30.2 zeigt die adäquate Stichprobenbeschreibung über den Verlauf einer Studie hinweg.

Gemäß dem Intention-to-treat-Prinzip (ITT) sollen möglichst alle Personen, die zu Studienbeginn eingeschlossen waren, in die Analyse eingehen (d. h. auch diejenigen, die die Intervention nicht vollständig erhalten bzw. abgebrochen haben). Kann dies nicht realisiert werden, muss die Art und Weise, wie mit unrealistischen Ausreißerwerten oder mit fehlenden Werten umgegangen wird, genau spezifiziert und begründet werden; die Möglichkeiten reichen hier von der einfachen Verwendung des jeweils letzten Wertes eines Teilnehmers, bevor er aus der Studie ausgeschieden ist (»last observation carried forward«, LOCF) bis hin zu komplexen mathematischen Modellierungen für die Schätzungen bei fehlenden Werten anhand der vorhandenen Daten.

Die auswertbare Analysestichprobe derjenigen, die den Versuch ordnungsgemäß durchlaufen haben, wird als »per protocol set« (oder auch: »valid cases«) bezeichnet.

Die vorab spezifizierten Auswertungen (z. B. einseitig versus zweiseitiger Test) sollten stets die Berechnung von geeigneten Vertrauensintervallen beinhalten. Ggf. muss bei Abhängigkeit verschiedener Messungen das Signifikanzniveau adjustiert werden. Die zentrale Auswertung (Wirksamkeitstest hinsichtlich der primären Variable) ist zu trennen von zusätzlichen Analysen (z. B. ob sich der Effekt in unterschiedlichen Subgruppen unterscheidet).

## Evaluation der Risiken und Nebenwirkungen (»safety and tolerability«)

### Wichtig

- Festlegung der Variablen und Grenzwerte
- Inzidenz unerwünschter Ereignisse (»adverse events«) und Teststärkenproblematik

Um die Unbedenklichkeit einer Intervention festzustellen, müssen unerwünschte Ereignisse zunächst über entsprechende Variablen zu möglichen Risiken und Nebenwirkungen (oder sogar Therapieschäden) und deren Berechnung definiert werden. In der Psychotherapieforschung wurde dieser Thematik bislang kaum systematisch Aufmerksamkeit geschenkt (vgl. Märtens & Petzold, 2002). Da die Inzidenz unerwünschter Ereignisse meist niedrig ist, ergibt sich hier eine besondere Teststärken-Problematik: Man benötigt große Stichproben, um seltene Ereignisse zu entdecken; somit ist es von besonderer Bedeutung, diese Aspekte auch in Phase-IV-Studien später weiter zu verfolgen.

## Abschließende Bewertung und Forschungsbericht

### Wichtig

- Umfassende deskriptive Auswertung
- Vergleichbarkeit (z. B. für spätere Datenaggregation) sichern

In [Tab. 30.2](#) ist eine Checkliste aufgeführt, welche Angaben in Forschungsberichten zu einer Interventionsstudie gemacht werden sollen (CONSORT Statement). Diese Standardisierung der Ergebnisdarstellung dient nicht zuletzt auch dazu, die Ergebnisse mehrerer Studien zu einem ähnlichen Untersuchungsgegenstand vergleichbar und aggregierbar zu machen (z. B. in Metaanalysen).

Im Anschluss an diese Beschreibung methodischer Prinzipien insbesondere von kontrollierten Wirksamkeitsstudien soll in der Studienbox auf S. 626 eine beispielhafte Studie der Phase III dargestellt werden, die hinsichtlich des methodischen Aufwands und ihrer hoch professionellen Durchführung als besonders anspruchsvoll gelten kann. Die bislang in diesem Kapitel entwickelte Analogie zur Pharmaprüfung setzt sich hier fort, da u. a. ein psychologisches Verfahren mit einem medikamentösen Ansatz verglichen wird.

## 30.2.4 Ausblick: »Patient focused therapy research« als Phase V der Therapieevaluation?

Wie bereits dargestellt, müssen sich die in kontrollierten Wirksamkeitsstudien als wirksam erwiesenen Verfahren in einer Phase IV unter klinischen Alltagsbedingungen bewähren (»effectiveness«), wobei in Studien zunehmend naturalistische Bedingungen hergestellt werden können. Man könnte sich dies folgendermaßen vorstellen: Eine manualisierte, in einer Wirksamkeitsstudie bewährte Intervention wird »aus dem Labor heraus« verbreitet (»dissemination«) und in einer oder mehreren repräsentativen klinischen Einrichtungen implementiert; dabei wird in einem nicht kontrollierten Prä-Post-Design für die ersten 350 Patienten evaluiert, ob sich die Ergebnisse aus den Wirksamkeitsstudien replizieren lassen. Kommt man zu vergleichbaren Ergebnissen und stellt fest, dass auch in dieser Feldstichprobe keine unvorhergesehenen Risiken und Nebenwirkungen auftreten, ist die Evaluation erfolgreich abgeschlossen, und die vier Phasen werden dann gegebenenfalls nach einer Weiterentwicklung des Verfahrens – z. B. wenn es theoretisch auf der Grundlage neuer wissenschaftlicher Faktenlage neu systematisiert und/oder mit neuen Komponenten versehen wird – erneut durchlaufen.

Doch während das Verfahren überarbeitet oder neue Verfahren entwickelt werden, braucht die Praxisforschung nicht stehen zu bleiben. Denn einerseits wird die Legitimation von Psychotherapie und anderen klinisch-psychologischen Interventionen weiter ausgebaut, wenn im Sinne einer »kontrollierten Praxis« Prozess und Ergebnis fortlaufend dokumentiert und evaluiert werden. Zudem kann eine kontinuierliche Sammlung von Praxisdaten dazu dienen, die angewandte Therapie auch im Einzelfall eigenständig »progressiv« weiterzuentwickeln, d. h. nicht nur bereits bestehendes Vorwissen im Stichprobendurchschnitt zu bestätigen.

Das Prinzip einer solchen »patient focused psychotherapy research« (Lutz, Martinovich & Howard, 2001; Lutz, 2002; Lambert, 2005) besteht darin, kontinuierlich Daten über Therapieverläufe zu sammeln und hinsichtlich der Patienten- oder Problemmerekmale Profile von erfolgreichen und nicht erfolgreichen Therapien herauszukristallisieren. Zeigt sich beispielsweise bei einem »typischen« Patienten mit Diagnose A, mit Schweregrad B und weiteren spezifizierten Randbedingungen im Verlauf ein Muster, das den bislang im System gespeicherten Verläufen ähnlich ist, die am Ende nicht erfolgreich waren, so könnte dem Therapeuten dies im Sinne eines Warnhinweises rückgemeldet werden (z. B.: »Achtung, eine ausbleibende Stimmungsaufhellung gemäß Fragebogen X nach bereits sechs Therapiesitzungen lässt bei diesem Patienten auch im Bereich Y

## Exkurs

## Effektstärken und klinische Signifikanz in Interventionsstudien

Effektstärken (ES) sind eine Art »gemeinsame Währung«, mit welcher auch Studien verglichen werden können, in denen sehr unterschiedliche Mess- und Auswertungsverfahren verwendet wurden. Mit Effektstärken wird das Ausmaß von Veränderungen und Gruppenunterschieden standardisiert. Die allgemeine Formel für Effektstärken entspricht dem Mittelwertsunterschied zwischen zwei Gruppen relativiert an der (Prä-)Streuung (die sich in kontrollierten Studien in den Gruppen nicht unterscheiden sollte):

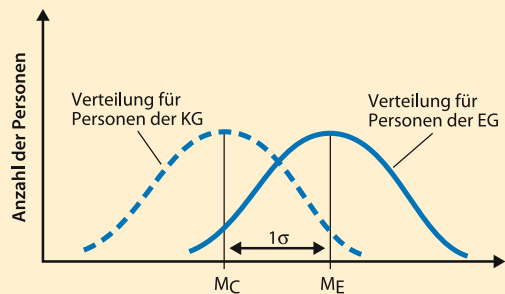
$$ES = \frac{M_a - M_b}{s}$$

In dieser Grundformel zur Effektstärkenberechnung ist ES die Effektstärke,  $M_a$  und  $M_b$  sind die Mittelwerte der verglichenen Gruppen und  $s$  bezeichnet die (Prä-)Streuung

Wie ist nun eine so ermittelte Effektstärke zu bewerten? Entscheidend sind sowohl die Mittelwertsdifferenz (z. B. der durchschnittliche Anstieg in einem Maß zur Lebenszufriedenheit) als auch die Streuung im Nenner, an der diese Mittelwertsdifferenz relativiert wird: Je kleiner die Streuung, desto stärker fällt ein Mittelwertsunterschied ins Gewicht. Das erklärt zu einem Teil auch, warum in der Praxis (heterogene Stichprobe, also große Streuung) meist niedrigere Effektstärken produziert werden als in einer kontrollierten Therapiestudie mit einer homogenen Patientenstichprobe (d. h. geringe Variabilität). Die Varianz durch Patientenselektion einzuschränken dient nicht nur dem »offiziellen« Ziel der Erhöhung interner Validität, sondern ist – bei nicht beliebig steigerbarer tatsächlicher Wirksamkeit – das einfachste Mittel zum Erhöhen von Effektstärken.

Eine Effektstärke von 1 – also die (positive) Veränderung des Patientenmittelwerts um eine Standardabweichung im Vergleich zu vor der Behandlung – bedeutet, dass es etwa 85% der Patienten nach der Behandlung besser geht als dem Durchschnittspatienten vorher.

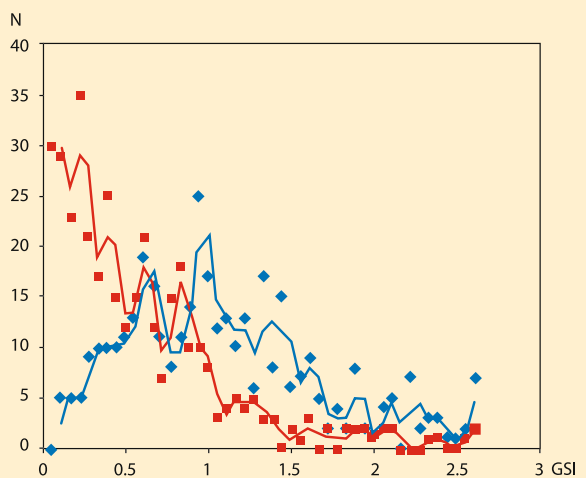
Prinzipiell können Effektstärken bei sehr großer Mittelwertsdifferenzen und sehr homogener Stichprobe, d. h. sehr kleiner Streuung, in ihrem Betrag unendlich groß werden – in der Realität sind Studien mit Effektstärken  $>3$  allerdings ausgesprochen selten. Cohen (1988) bezeichnet Effektstärken von 0,2 als geringe, Werte von 0,5 als mittlere und ab 0,8–1,0 als hohe Effektstärken. Rosenthal (1982, 1990) verweist auf die häufige Unterschätzung bei der Interpretation von Effektstärken. Er zeigt, dass eine Ef-



■ **Abb. 30.3.** Idealierte Darstellung einer Effektstärke von  $ES=1,0$

fektstärke von 0,85, wie sie von Smith, Glass und Miller (1980) in der ersten großen Metaanalyse zur Psychotherapie ermittelt wurde, ein äußerst bedeutungsvolles Ergebnis ist. Grawe et al. (1994) zeigen, dass bei einer Stichprobe von Untersuchungen die Effektstärken von insgesamt 111 Kontrollgruppen einen deutlichen Schwerpunkt um Null und leicht darüber haben. Regressionseffekte (Besserung ohne Therapie) sind längst nicht so stark, wie einst von Eysenck provokativ angenommen. Nähere Informationen zu Vor- und Nachteilen von Effektstärken, sowie zu verschiedenen Berechnungsvarianten finden sich z. B. bei Hartmann und Herzog (1995).

Die Bedeutung von Effektstärken wird in Lehrbüchern meist, wie in ■ **Abb. 30.3** dargestellt, idealisiert illustriert.



■ **Abb. 30.4.** Die Effektstärke von  $ES=1,0$  anhand realer Therapie-daten; Verteilung des SCL-90-Gesamtwertes (GSI) vor (blau) und nach (rot) Therapie





▼ Ist die Verteilung der Experimentalgruppe (EG) mit einem Mittelwert  $M_E$  gegenüber der Verteilung der Kontrollgruppe (KG) mit dem Mittelwert  $M_C$  um eine Standardabweichung verschoben, handelt es sich um eine Effektstärke von  $ES=1$ .

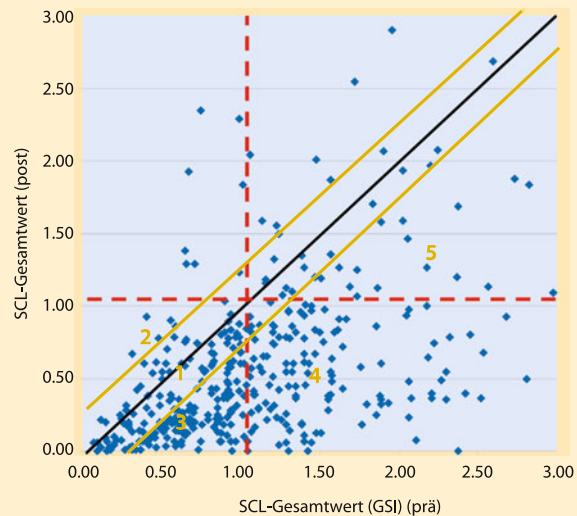
Reale Daten sehen aber meist anders aus. ■ Abb. 30.4 zeigt die Trendlinien der Verteilung des Gesamtwertes einer übergreifenden Symptomskala (SCL-90-R) anhand von  $N=376$  Patienten vor (blau) und ein Jahr nach Therapie (rot).

Die Ergebnisse sind erstens heterogener als in der idealisierten Darstellung und verteilen sich zweitens nicht normal. Insbesondere die Follow-up-Werte sind – durchaus erwünscht – schief verteilt, da eine niedrige Symptombelastung häufig und hohe Symptombelastung nur noch selten vorkommt. Die Effektstärke dieser empirisch ermittelten Verteilung liegt ebenfalls etwa bei 1.

Da bei der reinen Effektstärke individuelle Unterschiede herausgemittelt werden, ist es darüber hinaus wünschenswert, Angaben zur klinischen Signifikanz auf Einzelfallebene (z. B. Anteil erfolgreich behandelter Patienten) zu betrachten. Hierzu müssen wir zuerst bestimmen, welche Veränderungen überhaupt als bedeutsam erachtet werden kann (»reliable change«; vgl. Jacobson et al. 1999). Einen differenzierten und zugleich praktisch gut umsetzbaren Ansatz, bei dem sowohl die Größe der Veränderung als auch das Ausgangsniveau (Prä-Messung) berücksichtigt werden, wenn es um die Bestimmung von klinisch signifikanten Veränderungsraten geht, liefern Hiller et al. (2009).

In ■ Abb. 30.5 sind die Veränderungswerte der oben genannten 376 Patienten abgetragen, wobei fünf Gruppen unterschieden werden können.

Die schwarze Diagonale bezeichnet Nichtveränderung und der Bereich außerhalb der gelben Linien reliable Veränderungen. Gleich geblieben bzw. hinsichtlich des Gesamtwertes der Symptombelastung im SCL-90-R ohne reliable Veränderung ist somit die Gruppe 1 (dies entspricht 8% der Stichprobe). In Gruppe 2 (oberhalb der oberen gelben Linie) finden sich die Werte der Patienten, die sich



■ **Abb. 30.5.** Klinische Signifikanz einer Effektstärke von  $ES=1,0$  anhand beispielhafter realer Therapiedaten: Nichtveränderung (1), reliable Verbesserungen (3 – reliable Verbesserung, wobei bereits bei prä-Messung im Normbereich; 4 – reliable Verbesserung vom auffälligen in den Normbereich; 5 – reliable Verbesserung, allerdings auch bei post-Messung noch im auffälligen Bereich) und reliable Verschlechterungen (2) in der 1-Jahres-Katamnese ( $y$ -Achse) gegenüber dem Zeitpunkt vor Therapiebeginn ( $x$ -Achse)

reliabel verschlechtert haben (4%). Die roten gestrichelten Linien zeigen einen groben Normbereich für die SCL-90-R an (hier definiert als Mittelwert einer gesunden Population plus zwei Standardabweichungen). Gruppe 3 hat sich reliable verbessert, war aber schon vor der Therapie hinsichtlich der SCL-90-R in diesem Normbereich (28%). Gruppe 4 (44%) hat sich klinisch signifikant verbessert (d. h. aus einem auffälligen in einen unauffälligen Bereich verändert). Gruppe 5 hat sich ebenfalls reliable verbessert, liegt aber auch nach Therapieende noch im auffälligen Bereich (16%).

Eine solche Darstellung ist in Bezug auf die Beurteilung der Ergebnisse einer Intervention um einiges informationshaltiger als die basale Aussage »Die Effektstärke bezüglich der Symptom-Gesamtbelastung zwischen Therapiebeginn und 1-Jahres-Follow-up beträgt 1,0«.

einen ungünstigen Verlauf erwarten!«). Dies ist insbesondere vor dem Hintergrund von Bedeutung, dass Praktiker in der Regel therapeutische Misserfolge nicht sonderlich korrekt vorhersagen können. Solche Rückmeldeprozeduren können daher auf Grundlage einer ständig wachsenden Datenbasis, verbesserter mathematischer Modellbildung und einem gelungenen Umgang mit der Rückmeldung seitens der klinischen Praktiker, die an einem solchen System teilnehmen (z. B. Reflexion und Hypothesenbildung, Inan-

spruchnahme von Supervision), die Qualität der Behandlungen über die Zeit verbessern. Es können so Misserfolge und Therapieschäden minimiert, positive Effekte systematisch verstärkt oder Behandlungsdauern reduziert werden – auch wenn im therapeutischen Alltag nicht in jede einzelne Fallkonzeption das theoretische und empirische Gesamtverständnis einfließen kann.

Eine solche Praxisforschung kann die Qualität klinischer Interventionen nicht nur (»konservativ«) sichern,

## Studienbox

**Was wirkt besser: Medikamente oder Verhaltenstherapie? Eine methodisch besonders avancierte Psychotherapiestudie zur Behandlung der Panikstörung**

Barlow, Gorman, Shear und Woods (2000) berichten von einer großen multizentrischen Therapiestudie, in der die Behandlung der Panikstörung (nach DSM-III-R, mit und ohne »milde« Agoraphobie) in fünf verschiedenen Behandlungsbedingungen hinsichtlich ihrer Wirksamkeit verglichen wurden. Diese Behandlungsbedingungen waren:

1. Imipramin (trizyklisches Antidepressivum; verabreicht nach der Dosierung, die zu Studienbeginn »state-of-the-art« war, einschließlich regelmäßiger ausführlicher ärztlicher Kontakte),
2. ein manualisiertes kognitiv-verhaltenstherapeutisches Programm (CBT) zur Panikbehandlung,
3. eine medikamentöse Placebo-Kontrollbedingung,
4. Kombination Imipramin + CBT sowie
5. Kombination CBT + medikamentöses Placebo.

Diese Studie (N=324, Erhebungszeitraum: 1991–1998) war besonders aufwändig und sorgfältig geplant worden; einige Highlights seien im Folgenden genannt.

Da die Erforschung medikamentöser versus psychotherapeutischer Behandlungserfolge traditionell häufig parallel und von »verfeindeten Lagern« unabhängig voneinander erfolgte (ähnlich psychodynamischer und verhaltenstherapeutischer Therapieforschung), ist es als großer Verdienst von Barlow et al. zu betrachten, in dieser direkten Vergleichsstudie anerkannte Vertreter in verschiedenen Studienzentren »zusammenzubringen«, die sich jeweils mit ihrem (medikamentösen oder kognitiv-verhaltenstherapeutischen) Ansatz identifizierten. Damit wurde der sog. Allegiance-Effekt minimiert.



Psychotherapieforschung ist durchführungstechnisch extrem aufwändig. Das Foto zeigt vier von 70 Akten einer kontrollierten Therapievergleichsstudie

Ebenfalls hervorzuheben ist in diesem Zusammenhang, dass die Studie »blind« angelegt war derart, dass die Messung der Symptomatik bei den Patienten vor und nach der jeweiligen Behandlung sowie in einer Katamnese nach 6 Monaten von externen Evaluatoren durchgeführt wurde, die über die jeweilige Gruppenzugehörigkeit der Patienten nicht Bescheid wussten.

Für eine Psychotherapiestudie außergewöhnlich ist die zusätzliche Differenzierung nach Therapiebeendern und Therapieabbrechern: Es werden bei den Analysen *alle* Patienten berücksichtigt, die nach Abklärung von Einschluss- und Ausschlusskriterien die Studie begonnen haben. Dieser sog. Intention-to-treat-Ansatz ist bezüglich seiner Aussagekraft als viel stärker einzuschätzen als eine Studie, in der Ergebnisse lediglich für die »verfügbaren« Patienten dargestellt werden, die an einer Nachuntersuchung teilgenommen haben. Dieses Vorgehen ist aus zweierlei Gründen für Psychotherapiestudien eher selten: Erstens verschlingt es deutlich mehr Forschungsressourcen, auch alle Abbrecher nachzuuntersuchen und in die Analysen mit einzubeziehen – und andererseits fallen damit die Erfolgsquoten zwar realistischer, aber auch deutlich unvorteilhafter aus.

Folgende sieben Ergebnisse seien an dieser Stelle für die Studie von Barlow et al. zusammenfassend genannt:

1. Sowohl CBT als auch Imipramin und ebenso die Kombinationsbehandlungen wirken besser als (medikamentöses) Placebo alleine – allerdings nicht in allen Maßen und Untersuchungsphasen. So hält anfangs die Placebobedingung noch mit, verliert aber zunehmend an Bedeutung.
2. Im direkten Vergleich zeigten CBT und Imipramin vergleichbare Erfolgsquoten bei denjenigen, die die Therapie auch beendeten.
3. Imipramin hatte bei denjenigen, die die Therapie erfolgreich beendeten (»responder«), stärkere Effekte gegenüber CBT vorzuweisen.
4. Unter Berücksichtigung der Abbrecher (ITT-Analyse) war jedoch CBT insgesamt der Imipramin-Behandlung überlegen, was auf die erhöhte Abbrecherrate bei den Medikamenten zurückzuführen ist.
5. Die Erfolge der CBT waren stabiler, d. h. bei der medikamentösen Bedingung waren im Nachuntersuchungszeitraum mehr Rückfälle zu verzeichnen.
6. Die Kombination von CBT und Imipramin *reduzierte* die Stabilität des Therapieerfolgs gegenüber der Behandlung mit CBT alleine. Dies ist ein besonders be-



merkenwertes Resultat, das weiterer Untersuchung bedarf.

7. Die Erfolgsquoten der aktiven Behandlungsbedingungen unter Berücksichtigung aller Teilnehmer zum Follow-up-Zeitpunkt waren: ca. 30% (CBT), ca. 20% (Imipramin) und ca. 25% (CBT+Imipramin). Dies nimmt sich ganz anders aus als in vielen anderen Studien zur Behandlung von Panik mit Verhaltenstherapie (wo häufig 70–80% Behandlungserfolg berichtet werden) und verweist darauf, dass wir bei unseren angestrebten Erfolgsquoten bescheidener

bzw. realistischer werden müssen, wenn wir *alle* Patienten berücksichtigen wollen, die eine Behandlung beginnen.

Eine solche Studie erfordert einen außerordentlichen Aufwand. In der zentralen Publikation der Studie im hochrangigen »Journal of the American Medical Association« (JAMA), die übrigens im »Web of Science« mittlerweile über 300 Mal zitiert wurde (ein ausgesprochen hoher Wert), werden einschließlich der vier Autoren insgesamt 77 Personen genannt, die an der Studie mitgewirkt haben!

sondern konkret weiterentwickeln und optimieren. Sie stellt daher eine echte Ergänzung (»Phase V«?) der Phasen I–IV der Therapieevaluation dar.

### 30.3 Überblick zum Forschungsstand bei verschiedenen Interventionsverfahren

#### 30.3.1 Unterschiedliche Therapieformen – gleiche Wirkung?

Nach diesem in erster Linie methodischen Überblick sollen nun einige inhaltliche Befunde zur Wirksamkeit und Effektivität zur Sprache kommen. An der Frage der nachgewiesenen Wirksamkeit und der jeweiligen Überlegenheit verschiedener Therapieverfahren scheiden sich seit jeher die Geister. Zu Beginn dieses Kapitels wurden bereits die methodenkritischen Einwände von Eysenck (1952) erwähnt, die darauf abzielten, dass der saubere Nachweis der Überlegenheit von Psychotherapie gegenüber Spontanremission noch ausstehe. Später wurde die Auffassung vertreten, dass Psychotherapie zwar schon hilft (d. h. besser ist als Nichtbehandlung) – aber auf ganz unspezifische Weise: Alle therapeutischen Verfahren seien gleich wirksam. In der Tat erweist sich aus verschiedenen Gründen ein direkter Vergleich beispielsweise zwischen Verhaltenstherapie und Psychoanalyse in ein und derselben Studie als äußerst schwer durchzuführen, u. a. weil externe Einflüsse insbesondere bei unterschiedlich langen Behandlungen (unterschiedliche Therapiedosis) nicht ohne Weiteres über die untersuchten Gruppen hinweg konstant gehalten werden können. Es existieren in der psychologischen Forschung keine experimentellen Ansätze, die einen so langen – u. U. sogar viele Jahre andauernden – Zeitraum als unabhängige Variable untersuchen wie dies bei einer solchen Psychotherapiestudie nötig ist.

Unter anderem deswegen ist die Frage noch unzureichend geklärt, warum über viele Studien hinweg die unterschiedlichsten Ansätze und therapeutischen Herangehensweisen in unterschiedlichsten Settings doch oft zu ähnlichen Ergebnissen kamen. Teils wird dies als Artefakt einer zu wenig differenzierten Methodik angesehen (unzureichende Operationalisierung von Wirkfaktoren oder -mechanismen, zu heterogene Patientenpopulationen, zu wenig Messzeitpunkte etc.), teils als Auswirkung allgemeiner »unspezifischer« Wirkfaktoren (wie der Therapiebeziehung), teils als Teststärkeproblem, insbesondere dann, wenn man nicht ein Behandlungspaket insgesamt, sondern einzelne Komponenten testen möchte (weder Stichprobengröße noch Therapieeffekt können beliebig vergrößert werden). Auch der bereits erwähnte Allegiance-Effekt wirft Fragen auf (Caspar & Jacobi, 2003): Beachtet man die Varianz, die durch die Identifikation der Forschergruppe mit dem jeweils untersuchten Verfahren aufgeklärt wird, so könnte man zu dem Schluss kommen, dass ein (hinreichend umgrenztes) Verfahren lediglich von seinen eigenen Anhängern beforscht werden muss – es wird schon etwas Signifikantes dabei herauskommen. Oder, anders ausgedrückt: Was ist das Ergebnis einer Wirksamkeitsstudie wert, wenn es mit hoher Sicherheit bereits aus den Präferenzen der Untersucher vorausgesagt werden kann?

Bereits Kiesler (1966) spricht im Zusammenhang mit der scheinbaren Äquivalenz der Ergebnisse früherer Therapiestudien vom »Uniformitätsmythos«. Er kritisiert die oft übergeneralisierten Annahmen, die den »Therapietheorien erster Generation« und den daraus abgeleiteten Forschungsdesigns zugrunde liegen (► Abschn. 30.2.2). Dass alle Therapieformen vor allem unspezifisch wirken und letztlich äquivalent seien, muss aus heutiger Sicht bestritten werden, insbesondere angesichts der störungsspezifischen Perspektive, der Wirkfaktorenforschung sowie aufgrund der zunehmenden Nutzung und Weiterentwicklung psychologischer Grundlagenforschung

## Gut zu wissen

**Metaanalysen**

Um Studienergebnisse (im klinischen Bereich meist mit kleinen Stichproben) zu aggregieren, kann auf meta-analytische Verfahren zurückgegriffen werden, indem z. B. die Effektstärken oder andere standardisierte Ergebnismaße aus vergleichbaren Studien gemittelt werden. Hierbei ist zu beachten, dass die Ergebnisse der eingehenden Studien nicht unabhängig von ihrer Qualität zu sein scheinen: Cuijpers et al. (2010) zeigen in ihrer Metaanalyse zum Zusammenhang von Studienqualität und Effektstärken im Bereich der Depressionstherapie, dass Studien, die viele der in Abschn. 30.2.3 aufgeführten Anforderungen umgesetzt haben, deutlich niedrigere Effektstärken haben als diejenigen, die wenige dieser Anforderungen erfüllen. Eine laxer Handhabung von Qualitätskriterien scheint also die Ergebnisse zu verbessern – eine für Therapieforscher gefährliche Verführung? Aber auch die Analyse von Cuijpers et al. hat ihre Tücken, denn es ist z. T. fragwürdig, welche Kriterien die Autoren anlegen, um einer Studie »hohe Qualität« zu bescheinigen. Beispielsweise ordnen sie keine der über 100 Studien mit einer Wartelisten-Kontrollgruppe der »High-quality«-Gruppe zu und erwecken damit den Anschein, Psychotherapieforschung gegenüber der Pharmaforschung (in der Wartelisten-Kontrollgruppen unüblich sind) grundsätzlich abwerten zu wollen. In gewisser Weise erscheint die Arbeit damit als Replik auf die brisante und viel beachtete Metaanalyse von Kirsch et al. (2008), die wiederum den Nutzen der Pharmakotherapie bei leichten und mittelschweren Depressionen sehr kritisch bewertet.

Grundsätzlich besteht die Gefahr, bei der üblichen Integration mehrerer Studien in Metaanalysen Effekte zu verwischen, weil zu unterschiedliche Studien zusammengefasst werden (»Äpfel-und-Birnen-Problem«; engl. »apples and oranges«) oder weil eine Verzerrung dahingehend besteht, nur erfolgreiche Ergebnisse zu veröffentlichen (»Schubladen-Problem«, »publication bias«). Auch gute und umfangreich dokumentierte Studien bleiben oft für eine Metaanalyse schwer zugänglich (»graue Literatur«, z. B. Dissertationen, Projektberichte), ebenso wie nicht englisch verfasste Publikationen leicht unentdeckt bleiben können.

Trotz dieser möglichen Probleme können Metaanalysen eine hervorragende Forschungsstrategie darstellen. Rosenthal und Di Matteo (2001) stellen die Vorteile dieses Ansatzes in einer guten Übersichtsarbeit heraus.

Eine besonders lehrreiche Abhandlung zur Interpretation von Metaanalysen liefert Plath (1998): Hier werden am Beispiel der einflussreichen forschungsintegrativen Arbeit von Grawe et al. (1994) Schritt für Schritt Aussagekraft, Stärken und Grenzen von Metaanalysen dargelegt. Grundprinzip ist, ebenso wie bei anderen Forschungsformen, die Replizierbarkeit bzw. Nachvollziehbarkeit der Befunde. Ein umfassendes Regelwerk für systematische Übersichtsarbeiten und Metaanalysen liefert das »Prisma-Statement« (<http://www.prisma-statement.org>; Liberati et al. 2009); dort ist auch eine Checkliste derjenigen Punkte enthalten, die bei der Publikation der eigenen Studie zu beachten sind, damit die Ergebnisse in späteren Metaanalysen verwertet werden können.

In jüngster Zeit werden zunehmend brisante Metaanalysen publiziert, die in der Folge kontrovers diskutiert werden. Aussagen aus diesen Metaanalysen wie etwa »Langzeit-Psychoanalyse ist wirksam« (Leichsenring & Rabung, 2008, Kritik z. B. Rief & Hoffmann, 2009) oder »CBT ist unwirksam zur Rückfallprophylaxe bei Schizophrenie und affektiven Störungen« (Lynch et al. 2009; Kritik z. B. Lincoln, 2010) werden hierbei meist in Bezug auf Ein- und Ausschlusskriterien angegriffen. Dabei wird den Autoren oft unterstellt, selektiv Studien aus ihrer Metaanalyse ausgeschlossen zu haben, die zu einem Ergebnis kommen, das den Autoren nicht passe (z. B. Studie X, in der herauskommt, dass CBT die Rückfallraten bei bipolaren Störungen senkt, wurde nicht in die Metaanalyse aufgenommen, weil es sich nicht um »CBT« sondern um »Psychoedukation« gehandelt habe). Oder umgekehrt: Studie Y wurde in die Metaanalyse aufgenommen, obwohl es sich nicht um »Langzeit-Psychoanalyse«, sondern um »Psychodynamische Kurzzeittherapie« gehandelt habe. Derartige Debatten (z. B. in Form eines Briefwechsels, der im Nachgang zu einer Metaanalyse im selben Journal unter Moderation des Herausgebers publiziert wird) sind hoch spannend und lehrreich; sie machen aber auch deutlich, dass Metaanalysen nicht ungeprüft als Ultima Ratio zur Beurteilung von Interventionen herangezogen werden können.

(► Abschn. 30.4). Bevor darauf näher eingegangen wird, sollen jedoch noch – unter dem Vorbehalt insgesamt recht unspezifischer Aussagen und der Probleme von Metaanalysen – einige nicht zuletzt berufspolitisch wichtige Ergebnisse der »klassischen« Psychotherapie-Outcome-For-

schung der 1950er bis 1980er Jahre genannt werden. Gemäß Grawe (1995) würde dies also den »konservativen« Forschungsstand zu Behandlungsverfahren betreffen, deren Strategien und Interventionen auf »Theorien erster Generation« beruhen.

**Studienbox****Einige zentrale Befunde aus Therapiestudien der 1950er bis 1980er Jahre (Grawe, Donati & Bernauer, 1994)**

Eine besonders umfassende, aber auch kontrovers und viel-diskutierte Metaanalyse ist die Arbeit von Grawe et al. (1994; »Psychotherapie im Wandel: Von der Konfession zur Profession«). Hier wurde der bislang einmalige Versuch unternommen, die ganze Grundgesamtheit der bis dahin publizierten Psychotherapiestudien zu integrieren. Kritiker warfen den Autoren – z. T. in unsachlicher Weise – vor, in ihrer Analyse verhaltenstherapeutische Verfahren zu begünstigen; außerdem wurde bemängelt, dass die einbezogenen Studien zwar zunächst hinsichtlich ihres Qualitätsprofils beurteilt wurden, diese Qualitätsprofile aber bei den Endauswertungen dann nur unzureichend berücksichtigt worden seien. Dem unbenommen stellt die Arbeit – nicht zuletzt aufgrund der Fokussierung auf empirische Ergebnisse (die die Versorgungsrealität weitgehend nicht erreicht hatten) und dem Hinterfragen von Glaubensrichtungen – einen großen Wert dar. Einige zentrale Befunde sind:

- Die Auswirkungen von Psychotherapie sind i. Allg. positiv (auch wenn der Forschungsstand zu »Risiken und Nebenwirkungen« von Psychotherapie noch immer unbefriedigend ist).
- Die Effekte von Psychotherapie sind deutlich höher als die durch spontane Remission und können auch im Vergleich mit medizinischen Interventionen als durchaus hoch angesehen werden.
- Insgesamt haben Verfahren aus dem kognitiv-behavioralen Spektrum die breiteste empirische Basis vor-

zuweisen. Dies hängt sicherlich stark mit dem Selbstverständnis der Verhaltenstherapie zusammen, denn empirische Fundierung und stetige Weiterentwicklung zählen zu ihren Grundprinzipien (Margraf & Lieb, 1994). Auch tiefenpsychologische (Kurzzeit-)Therapie und Gesprächstherapie können wissenschaftliche Belege ihrer Wirksamkeit anführen. Daraus allgemeine Schlüsse zu ziehen, etwa dass Psychodrama, systemische und Gestalttherapie unwirksam seien oder dass tiefenpsychologische Verfahren als Richtlinienverfahren für alle Störungsbereiche indiziert sind, ist angesichts des derzeitigen Forschungsstandes noch nicht angemessen.

Die umfangreiche Sammlung von Psychotherapiestudien bei Grawe et al. (1994) diente – z. B. mit über 2300 Befunden bis 1992 zu Prozessmerkmalen und Therapieergebnis – auch als eine Basis dafür, die Wirkfaktorenforschung voranzutreiben (► Kap. Abschn. 30.4).

Eine Auflistung von über 250 weiteren bzw. neueren Metaanalysen aus dem Bereich klinisch-psychologischer Interventionen findet sich auf der Webseite von William R. Shadish (<http://faculty.ucmerced.edu/wshadish/psychotherapy.htm>). Eine gute Quelle für die Suche aktueller Evidenzen ist auch die Cochrane Library (<http://www.thecochranelibrary.com>; s. Abschn. 30.3.2)

### 30.3.2 Empirically Supported Treatments (EST)

Dass Analogien bezüglich der Entwicklung und Bewertung klinisch-psychologischer Interventionen einerseits und Psychopharmaka andererseits bestehen, wurde bereits eingangs deutlich. Die Entwicklung sog. »empirically supported treatments« (EST) ist davon stark geprägt und hat eine wichtige übergreifende Debatte in der Psychotherapieforschung initiiert.

Zum Hintergrund: Angesichts der stark ungleichen Budgets von Psychotherapie- und Pharmaforschung beschloss die American Psychological Association zu Beginn der 1990er Jahre eine »Gegenoffensive«. Die klinisch überprüfte Leistungsfähigkeit psychologischer Verfahren sollte analog der »evidence based medicine« (vgl. Sackett et al. 1996; eine stets wachsende Sammlung an Evidenz im ganzen medizinischen Bereich findet sich in der Cochrane Library; <http://www.thecochranelibrary.com>) dokumentiert und herausgestellt werden. Kriterium für »empirische

Fundierung« ist danach die Existenz von Studien, die neben der Randomisierung der Patientenzuweisung (oder einer Serie kontrollierter experimenteller Einzelfallstudien) eine Manualisierung des Vorgehens und eine homogene Patientenauswahl nach klaren diagnostischen Kriterien verwenden. Zudem müssen Ergebnisse durch eine zweite Forschergruppe bestätigt werden. 1995 folgte eine Liste »offiziell anerkannter« Verfahren (dt. Hahlweg, 1995), die seitdem mehrfach ergänzt wurde (vgl. auch Nathan & Gorman, 2002, Fonagy & Roth, 2004).

Somit steht nicht mehr die (relative) Leistungsfähigkeit einer ganzen Therapierichtung im Zentrum neuerer Entwicklungen, sondern die Therapie verschiedener Störungen mit ihren jeweiligen Besonderheiten. Behandlungsprogramme und -manuale werden störungsspezifisch entwickelt und evaluiert. Die Behandlung einer Blut- und Spritzenphobie unterscheidet sich von der einer Agoraphobie, beide von einem Programm für Sozialphobiker – und alle Vorgehensweisen gehen über therapieschulenspezifische Standardmethoden hinaus. Diese Entwicklung veranlasste

## Gut zu wissen

**Die Debatte um Empirically Supported Treatments (EST) 1995 bis 2005**

Nach der (Weiter-)Entwicklung der ersten EST-Konzepte (Chambless & Hollon, 1998; Kendall, 1998) ist eine ständige Diskussion zu verzeichnen, in der um die Definition von »empirischer Fundierung« bzw. um die Interpretation der Befundlage gestritten wird. Kritik am EST-Prinzip betraf meist die externe Validität von Ergebnissen, die durch randomisierte kontrollierte Wirksamkeitsstudien gewonnen werden: Praxisbedingungen sind bei diesem experimentellen Vorgehen schwer zu standardisieren und Langzeittherapien praktisch gar nicht durch Wirksamkeitsstudien zu untersuchen. Einer der fundamentalen Unterschiede zwischen einer Behandlung im Rahmen einer Wirksamkeitsstudie und einer Behandlung im klinischen Alltag ist z. B., dass in Ersterer »allgemeine Wirkfaktoren« streng kontrolliert werden müssen, um Gruppenunterschiede auf die spezifische Behandlung zurückführen zu können – eine Strategie, die für die Routinebehandlung sicher kontraproduktiv wäre (► Abschn. 30.4.1). Zudem erscheint es problematisch, dass die in Wirksamkeitsstudien notwendigen Ein- und Ausschlusskriterien in der Vergangenheit oft dazu führten, dass in der Regel nur etwa 40–50% der verfügbaren Patienten in die Studie aufgenommen werden.

Skeptiker befürchten einen ungerechtfertigten Ausschluss therapeutischer Verfahren aus Praxis und Forschung und durch die Manualisierung eine Entwicklung in Richtung »Kochbuch-« oder »Malen-nach-Zahlen-Psychotherapie«.

Ferner wird darauf hingewiesen, dass

- untersuchte Bündel von Verfahren mit dem gleichen Label (z. B. »kognitiv-behaviorale Therapien«) über verschiedene Settings und Studien oft schwer vergleichbar sind; Malik et al. (2003) zeigen etwa anhand einer Multizenterstudie mit vielen verschiedenen Behandlungsbedingungen, dass sich »kognitive Verfahren« durchaus hinsichtlich mancher Prozessmaße voneinander unterscheiden und gleichzeitig nicht-kognitiven Verfahren ähneln können;
- nur bei einem Teil unselegierter Patienten eine bestimmte Störung im Vordergrund steht und dass es schwierig sein wird, für seltenere Störungen und komorbide Kombinationen von Störungen mehrfach evaluierte Therapieprogramme zu entwickeln;
- einige der Grundannahmen von Wirksamkeitsstudien nicht allgemein gelten (z. B. dass jegliche Psychopathologie mit speziellen, meist kurzen Interven-

tionen veränderbar sei; dass stets ein primäres Problem identifiziert und, ohne Berücksichtigung weiterer Bereiche wie etwa Persönlichkeitsmerkmale bearbeitet werden kann; oder dass Wirksamkeitsstudien den einzigen Goldstandard zur Bewertung von Therapieverfahren darstellen);

- eine einseitig störungsorientierte Sicht eine Berücksichtigung von Ressourcen (► Kap. 21) erschweren kann, deren Bedeutung empirisch immer evidenter wird, und dass deshalb – so bedeutsam die Entwicklung störungsspezifischer Vorgehensweisen ist – eine zu starke Orientierung daran bedeutet, weitere Potenziale zu verschenken (► Abschn. 30.4),
- die Langzeiteffekte von EST – zumindest bei bestimmten Störungen mit fluktuierendem Verlauf wie z. B. Depressionen (vgl. Kopta et al. 1994) – nicht so optimistisch beurteilt werden dürfen wie ursprünglich angenommen. So kamen etwa Shea et al. (1992) in einer der größten Psychotherapiestudien, die jemals durchgeführt wurde (National Institute of Mental Health Treatment of Depression Collaborative Research Program), zu dem Schluss, dass eine 16-wöchige Behandlung mit kognitiv-behavioraler, interpersoneller oder medikamentöser Therapie für die meisten depressiven Patienten nicht ausreicht, um langfristige Effekte zu erzielen.

Oft greift die Kritik dabei zu kurz. So ist etwa das Argument nicht wahr, in Studien, die für die Bewertung von EST herangezogen werden, handle es sich ausschließlich um hoch selegierte Patienten, die der Versorgungspraxis nicht entsprechen: Viele Wirksamkeitsstudien hatten durchaus nicht jegliche Komorbidität ausgeschlossen, und ein erfolgreicher »Transport« von in experimentellen Studien als wirksam erachteten Interventionen in den Versorgungsalltag wurde mehrfach belegt (z. B. Cukrowicz et al. 2005; Hahlweg et al. 2001; Hiller et al. 2003; Shadish et al. 2000; Wade, Treat & Stuart, 1998; Zielke, 1993). Einen besonders interessanten Ansatz lieferten in diesem Zusammenhang Franklin et al. (2000): Dort wurden in einer Effektivitätsstudie an einem Studienzentrum, in dem eine Wirksamkeitsstudie zur Behandlung von Zwangsstörungen durchgeführt wurde, diejenigen Patienten behandelt, die aufgrund von Ausschlusskriterien (z. B. zusätzliche Diagnose einer Substanzstörung oder gleichzeitiger Medikation) nicht in den Wirksamkeitsstudien aufgenommen worden waren. Auch hier kam es zu vergleichbaren Ergebnissen, was die These einer grund-

▼  
sätzlichen »Nichtübertragbarkeit von Laborergebnissen« in Frage stellt.

Ferner ist der Vorwurf, der EST-Ansatz beschränke sich auf einen dichotomen »Ist-auf-der-Liste«- versus »Ist-nicht-auf-der-Liste«-Ansatz nicht mehr allgemein gerechtfertigt, da zunehmend versucht wird, Studien eher hinsichtlich theoretisch fundierter Grundannahmen über die Natur psychologischer Probleme und die Natur therapeutischer Veränderungsmechanismen auszurichten (»empirically supported principles of change«; z. B. Rosen & Davison, 2003).

Wer sich für neuere Entwicklungen dieser Diskussion interessiert, dem sei eine Debatte – also eine Serie von aufeinander Bezug nehmenden Artikeln – im »Psychological Bulletin« um die anspruchsvolle Überblicksarbeit von Westen, Nowotny und Thompson-Brenner (2004) empfohlen. Nutzen Sie dafür Ihren regional verfügbaren Online-Zugriff auf Fachzeitschriften, in diesem Falle auf acht Artikel von 2004/2005 im »Psychological Bulletin« 130(4)–131(3); oder bekommen Sie heraus, wie man an einen solchen Zugriff herankommt, auch wenn die eigene Universität keinen hat!

Barlow (2004) zu empfehlen, nunmehr nicht mehr von »klassischer Psychotherapie« (»generic psychotherapy«), sondern von psychologischen Behandlungen bzw. Interventionen (»psychological treatments«) zu sprechen. Er zeigt dabei auf, dass die Überlegenheit solcher psychologischer Verfahren z. B. gegenüber medikamentöser oder anderer medizinischer Interventionen um so deutlicher wird, je spezifischer gestörte Funktionen oder Funktionsmuster behandelt werden. Barlow verweist hier aber darauf, dass in den USA psychologische Interventionen oft von Hausärzten und medizinischem Personal ohne genuin psychologische Ausbildung durchgeführt werden und fordert, dass hier mehr Psychologen zum Einsatz kommen sollten. Dass insbesondere bei schwereren Fällen die Interventionen oft nicht so erfolgreich sind, wenn sie von nicht psychologischen, klinisch weniger geschulten Personen angewandt werden, könnte daran liegen, dass trotz Störungsspezifität weitere allgemeine Faktoren für die Umsetzung der Maßnahme eine Rolle spielen (► Abschn. 30.4.1).

Um derartige spezifische Forschung voranzutreiben, genügt der reine Wirksamkeitsnachweis nicht. In entsprechenden Studien sollte daher stets versucht werden, die Natur der therapeutischen Veränderung mit zu untersuchen. Entsprechende Forschungsdesigns könnten z. B. zusätzliche aktive Interventionsbedingungen mit mehr oder weniger potenziell wirkenden Elementen enthalten (z. B. »dismanteling design«, »component control design«), um einzelne Komponenten hinsichtlich ihrer Wirksamkeit beurteilen zu können (Borkovec & Castonguay, 1998). Eine solche »Dismanteling«-Strategie ist allerdings nur dann nützlich, wenn eher kurze, gut umschriebene und strikt manualisierte spezifische Behandlungsprozeduren untersucht werden; ihr Einsatz für die Erforschung umfassender Behandlungspakete, die sich nicht einfach additiv aus einzelnen Behandlungsmodulen zusammensetzen, wäre eher problematisch. Dass der Einsatz solcher Methodiken vor dem Hintergrund eines Wirksamkeitsansatzes nicht für alle

Verfahren und Behandlungspakete sinnvoll ist, gehört zu den wichtigsten Kritikpunkten im Rahmen der EST-Debatte (► Kasten S. 630f). Rosen und Davison (2003) oder Westen, Nowotny und Thomson-Brenner (2004) betonen ebenfalls, dass statt einer Liste empirisch gut fundierter Therapieverfahren eher empirisch fundierte therapeutische Prinzipien erarbeitet werden sollten.

### 30.4 Die Erforschung von Wirkfaktoren und Wirkmechanismen

Die bisherigen Ausführungen beschäftigten sich vornehmlich mit der Frage, ob eine neue oder weiterentwickelte Intervention wirksam ist. Hierfür gibt es wie gezeigt eine Fülle von relativ klar umschriebenen Konventionen und Richtlinien. Forschungsansätze dahingehend, wie Studien beschaffen sein müssen, um Beiträge zum »Wie« dieser Wirkung zu liefern, sind dagegen aufgrund der Fülle von Möglichkeiten nicht so einfach darzustellen. So gibt es z. B. allein zum Bereich der »therapeutischen Beziehung« als Wirkfaktor ganze Kompendien zur bisherigen Forschung (Norcross, 2003). Daher soll hier eher exemplarisch auf einige der zentralen Aspekte eingegangen werden. Einen umfassenden Versuch, auf der Basis der Vielzahl bereits vorliegenden Erkenntnisse zur Wirksamkeit verschiedener Prozessvariablen, die bei klinisch-psychologischen Interventionen von Bedeutung sind, allgemeine Wirkfaktoren zu extrahieren, liefert Grawe (1995, 1997, 1998).

Um die komplexen Prozesse in einer Psychotherapie auch im Verlauf adäquat abbilden zu können, reichen zudem eine Prä- und eine Post-Messung keinesfalls aus, sondern es sollten regelmäßig – möglichst nach jeder Therapie-sitzung – Zwischenmessungen vorgenommen werden; ein Beispiel für einen solchen relativ umfangreichen Stundenbogen, der nicht nur Ergebnisse, sondern auch die Realisie-

rung allgemeiner Wirkfaktoren misst, liefern Flückiger et al. (2010).

### 30.4.1 »Allgemeine« und »unspezifische« Faktoren: Psychologische Intervention = Placebo?

Dass in Therapiestudien viele unterschiedliche Verfahren zu ähnlichen Ergebnissen kamen, wurde zunächst damit erklärt, dass allgemeine, unspezifische Wirkfaktoren zum Tragen kommen. In seinem Buch »Persuasion and Healing« entwickelte Jerome Frank (1961) ein Modell, nachdem die Veränderung von Erwartungen als zentraler Wirkmechanismus angenommen wird. Jede Art von Psychotherapie sei im Prinzip geeignet, bei einem demoralisierten, Hilfe suchenden Patienten Hoffnung auf Besserung zu induzieren. Folgende Wirkfaktoren spielten in diesem Zusammenhang eine Rolle:

1. eine als Therapiebeziehung definierte Beziehung zwischen einem Hilfeempfänger und einem Hilfegeber mit einer speziellen Ausbildung, die Letzteren für diese Tätigkeit qualifiziert und in den Augen des Patienten kompetent erscheinen lässt,
2. ein formalisiertes Behandlungsangebot in einem gewissen institutionellen Rahmen (Klinik, Praxis, Ambulanz etc.), wobei das Setting selbst nach Frank schon zur Erwartung beiträgt, kompetente Hilfe zu bekommen,
3. ein bestimmtes Behandlungsrational, auf dessen Grundlage dem Patienten sein Zustand erklärt und eine Behandlungsstrategie abgeleitet wird,
4. die Durchführung eines mit dem Behandlungsrational konsistenten Behandlungsrituals oder Behandlungsvorgehens.

Entscheidend für die Wirkung sei, wie glaubwürdig im konkreten Einzelfall diese vier Bedingungen realisiert werden. Durch die Induktion positiver Erwartungen wird demnach ein sich selbst aufrechterhaltender Prozess in Gang gebracht (Zuversicht → mehr Energie, an Dinge heranzugehen → erste kleine Erfolgserlebnisse → erhöhte Kooperationsbereitschaft und Engagement → weitere Erfolge → mehr Zuversicht etc.), der wie eine sich selbst erfüllende Prophezeiung schließlich dazu führt, dass beim Patienten irgendwann tatsächlich eine positive Änderung eingetreten ist.

#### Wichtig

In der Tat zählt die wahrgenommene Kompetenz und Glaubwürdigkeit des Therapeuten zu den empirisch gut abgesicherten Prozessmerkmalen, die einen positiven Einfluss auf das Therapieergebnis haben.

Geht es also bei klinisch-psychologischen Interventionen nur um die Durchführung eines möglichst glaubwürdigen Rituals, und sollte man also eher von »Placebo-« als von »echten« Effekten sprechen? Placeboeffekte, also die Wirkung von »Scheinmedikamenten« ohne »echten« Wirkmechanismus können u. a. durch Erwartungsinduktion und durch spezielle Konditionierungsprozesse erklärt werden (Stewart-Williams, 2004). Grawe (1998, S. 26ff.) problematisiert die Anwendung der Placebobegrifflichkeit auf psychotherapeutische Verfahren, da einerseits meist unklar bleibt, um was es sich bei der Placebowirkung überhaupt handelt und andererseits, weil dadurch das Schulenden im Rahmen von Therapietheorien älterer Generation gefördert wird. Einflüsse, die nicht in der speziellen Theorie der jeweiligen Therapieschule vorgesehen sind, werden zum Placebo oder zu unspezifischen Wirkfaktoren erklärt, wobei diese Begriffe etwas weniger Wichtiges suggerieren – als »wirklich wichtige« Einflussfaktoren werden dagegen die spezifischen Wirkfaktoren der jeweiligen Therapietheorie postuliert. Warum sollte aber der durch die Induktion positiver Veränderungserwartungen angestoßene positive Rückkopplungsprozess zu einem unspezifischen oder in irgendeiner Hinsicht zweitrangigen Wirkmechanismus erklärt werden, wenn er zu so bedeutenden Verbesserungen führt, wie es den Anschein hat? Grawe stellt heraus, dass die systematische Nutzung von Erwartungsänderung bislang vernachlässigt wurde und illustriert an zwei Beispielen, wie fruchtbar diese Perspektive für die Interventionsforschung sein kann (► Kasten »Gut zu wissen«).

Erwartungsinduktion ist natürlich nicht der einzige Wirkmechanismus, der Besserungen schon bald nach Beginn einer Therapie herbeiführen kann; andere Wirkfaktoren können hinzukommen und eine eigenständige Rolle spielen. Aber die Voraussetzungen für das Wirksamwerden solcher weiterer Einflüsse werden besser, wenn durch eine Induktion positiver Erwartungen der Boden dafür bereitet wurde (Grawe, 1998, S. 31).

### 30.4.2 Sollten klinisch-psychologische Interventionen standardisiert oder individualisiert durchgeführt werden?

Die Erforschung von Wirkfaktoren und Wirkmechanismen, die der Wirkung klinisch-psychologischer Interventionen zugrunde liegen, schließt also die Art und Weise der »Verabreichung« der Intervention ein. Obwohl wir z. B. wissen, dass die Exposition mit angstauslösenden Situationen bei der Therapie von Angststörungen eine zentrale Rolle spielt, schließt sich daran immer noch die Frage an,



**Gut zu wissen****Beispiele für die Wirkung der Erwartungsänderung**

Grawe (1998) verweist zunächst auf die Methode der systematischen Desensibilisierung, von der schon erwähnt wurde, dass ihre Wirkungsweise nicht bzw. nur unzureichend mit den ursprünglichen Annahmen ihres Entwicklers erklärt werden kann. Bei der systematischen Desensibilisierung werden die vier von Frank (1961) postulierten Wirkfaktoren in besonderem Maße verwirklicht. Ein Therapeut, der eine schulmäßige systematische Desensibilisierung durchführt, zelebriert nach heutigem Erkenntnisstand tatsächlich ein Ritual, dessen Wirksamkeit wesentlich in seiner Glaubwürdigkeit, aber nicht in seinen einzelnen prozeduralen Komponenten begründet liegt. Damit ist nicht der Nutzen der Methode in Frage gestellt, sondern durch diese zusätzlichen Erklärungsansätze »neuerer Generation« wurde ein neuer Erklärungshorizont eröffnet, der über diese Therapiemethode hinaus neue Fragestellungen und praktische Schlussfolgerungen nahe legt.

Als zweites Beispiel für die Notwendigkeit, Erwartungseffekte bei der Entwicklung und Beurteilung kli-

nisch-psychologischer Interventionen zu berücksichtigen, nennt Grawe die aufschlussreiche Studie von Southworth und Kirsch (1988). Dort erhielten zwei Gruppen von Agoraphobikern die Instruktion, sich so weit von zu Hause fortzubewegen, bis sie Angst bekämen und dann umzudrehen. Dies sollte im Zeitraum von 2 Wochen 10-mal wiederholt werden. Für die eine Gruppe wurde dies als Beginn der Behandlung dargestellt, für die andere Gruppe als diagnostische Maßnahme zur genaueren Erfassung der Angst, und die »eigentliche Behandlung« würde erst nach diesen 2 Wochen beginnen. In beiden Gruppen nahm die Angst gegenüber einer unbehandelten Kontrollgruppe ab, aber in der ersten Gruppe mit therapeutischer Instruktion war dieser Effekt viel stärker ausgeprägt als bei der Gruppe, die davon ausging, dass die Therapie noch gar nicht begonnen hat. Bei der Auswertung war die jeweilige Zeitdauer der Reizkonfrontation kontrolliert worden, d. h. es war sichergestellt, dass beide Gruppen der phobischen Situation gleich lange ausgesetzt waren. Die Erwartung einer wirksamen Behandlung spielt also auch bei der In-vivo-Reizkonfrontation eine wichtige Rolle.

wie diese Maßnahme optimal zu realisieren bzw. in eine übergreifende Behandlungsstrategie zu integrieren ist (► Studienbox).

### 30.4.3 Beispiel für die systematische Weiterentwicklung einer Therapie der posttraumatischen Belastungsstörung

Als Beispiel für ein Forschungsprogramm, bei dem eine Behandlungsmaßnahme durch die Nutzung von experimentellen Befunden kognitionspsychologischer Natur systematisch weiterentwickelt wurde, sei hier die kognitive Therapie der posttraumatischen Belastungsstörung genannt (► Kap. 45). Ehlers, Clark, Hackmann, McManus und Fennell (2005) stellen in ihrer Übersichtsarbeit zur (Weiter-)Entwicklung des Therapieansatzes zuerst wichtige Ziele für die Traumabehandlung vor, die aus der bisherigen Forschung abgeleitet werden können:

1. Modifiziere eine extrem negative Bewertung des Traumas und seiner Folgen! Solche bis ins exzessive gehenden negativen Bewertungen (»appraisals«) des traumatischen Ereignisses können insbesondere über die Bedeutung von »hot spots« (unangenehmste Momente, die im Traumagedächtnis gespeichert sind) identifiziert werden.

2. Reduziere Symptome des Wiedererlebens durch Elaboration des Traumagedächtnisses und Diskrimination von Auslösern (Trigger)! Therapeutisch zentral ist hier das Durcharbeiten des Traumas (Rekonstruktion des Hergangs, Üben des Trennens zwischen »damals« und »heute«, wenn in Flashbacks das Trauma oder Teile davon subjektiv hoch realistisch wiedererlebt werden).
3. Unterbinde dysfunktionale Verhaltensweisen und kognitive Strategien! Insbesondere ausgeprägtes Vermeidungs- und übersteigertes Sicherheitsverhalten (offenes und gedankliches) trägt bei der posttraumatischen Belastungsstörung ebenso wie bei anderen Störungsbereichen zur Aufrechterhaltung der Problematik bei.

Vor diesem Hintergrund stellen Ehlers et al. (2005) dann heraus, dass das von ihnen entwickelte kognitive Behandlungsmodell gegenüber den bislang etablierten Ansätzen, die durchaus an ähnlichen Punkten ansetzen, Besonderheiten aufweist. Erstens wird die Technik des Wiedererlebens in sensu, die in fast allen Ansätzen vorkommt, verkürzt und unter einem neuen Blickwinkel eingeordnet (weniger zur Habituation an schmerzhaftes Erinnerungen als zur Identifikation von Hot Spots angewendet, die dann mit kognitiven Umstrukturierungsmaßnahmen bearbeitet werden). Als Vorteil kann hier eine bessere Verträglich-

## Studienbox

**Gekettete Kontrollgruppen**

Im Bochumer Angsttherapieprojekt wurde an einer Stichprobe phobischer Patienten die Nützlichkeit einer individuellen Problemanalyse und Therapieplanung mit einem Standardvorgehen (Reizkonfrontation in vivo kombiniert mit einem Selbstverbalisationstraining) als Kontrollbedingung verglichen (Schulte, 1993). Zeigten sich nun Unterschiede im Gesamterfolg, so konnten diese jedoch nicht eindeutig auf das Merkmal individualisiertes versus standardisiertes Vorgehen zurückgeführt werden: Als Ergebnis individueller Problemanalysen können verschiedene, mehr oder weniger effektive Methoden zum Einsatz kommen, die für unterschiedliche Ergebnisse verantwortlich sein könnten. Ein einfacher Vergleich der Bedingungen Problemanalyse und Standardtherapie kann darüber keine Auskunft geben.

Wie lösten Schulte und Mitarbeiter das Problem? Eine weitere, sog. gekettete Kontrollgruppe wurde gebildet. Patienten dieser Gruppe erhielten nach Zufall als Behandlungsmanuale die individuellen Therapiepläne der Problemanalysegruppe. Sie wurden also mit den gleichen Methoden behandelt – nur dass diese nicht für sie geplant wurden, sondern ihnen zufällig zugeordnet wurden. Damit ist eine Trennung der Effekte individueller Planung und Methode möglich, denn Unterschiede zwischen der Problemanalysegruppe und ihren »Zwillings-therapien« in der geketteten Kontrollgruppe können nur

aufgrund unterschiedlicher Passung der jeweils eingesetzten Methoden zustande kommen.

Insgesamt erwies sich in der genannten Studie die Standardbehandlung den beiden anderen Gruppen überlegen, und zwar stabil in Bezug auf das gewählte Erfolgsmaß, die Katamnesedauer und die Erfahrung des Therapeuten. Die Methode (Reizkonfrontation) trug also stärker zum Therapieerfolg bei als die Individualisierung bzw. die Passung. Es schien den Therapeuten nicht nur nicht zu gelingen, die Standardbehandlung durch individuelle Problemanalyse noch weiter zu verbessern, sondern die individualisierten Therapien schnitten dem Standardvorgehen gegenüber sogar schlechter ab. Die Autoren interpretieren diesen kontraintuitiven Befund dahingehend, dass die klare Vermittlung und Umsetzung eines etablierten Standardvorgehens – zumindest bei Phobikern – möglicherweise sowohl für Therapeuten als auch für Patienten insbesondere zu Therapiebeginn zielführender ist, als eine weniger stringente und potenziell sogar in sich widersprüchliche individuelle Fallkonzeption (► Abschn. 30.4.1). Es könnte zudem sein, dass eine Therapieplanung aufgrund einer individuellen Problemanalyse *doch* effektiver sein kann als die Standardbedingung – und dass in dieser Studie möglicherweise die individuellen Problemanalysen lediglich nicht die adäquaten bzw. bestmöglichen waren (Schulte, 1993).

In wie weit sich dieser Befund auch auf andere Interventionen und Populationen generalisieren lässt, ist noch unklar, da Studiendesigns dieser Art bislang selten sind.

lichkeit (und damit niedrigere Abbruchraten) erwartet werden als bei Verfahren, bei denen die umfassende, lang anhaltende Konfrontation mit dem Trauma im Mittelpunkt steht.

Auch beim Durchbrechen von Vermeidungsverhalten (z. B. Aufsuchen des Traumaortes) wird der Fokus nicht auf das Durchhalten und Habituiere, sondern auf das Bearbeiten spezifischer kognitiver Fehler (z. B. Übergeneralisierung von Gefahren) gerichtet. Insbesondere wird dabei systematisch geübt, das Gefühl »Es ist genau so, als ob es wieder geschehe ...« zu reduzieren, indem unter Einfluss der angstbesetzten Situation herausgearbeitet wird, welche Unterschiede zwischen »damals« und dem »Hier und Jetzt« bestehen. Dabei werden spezielle neue Techniken eingesetzt, um eine solche Stimulusdiskrimination zu erreichen (d. h., um zu verhindern, dass mit dem Trauma verbundene, aber eigentlich neutrale Reize erneut als Auslöser von Symptomen des Wiedererlebens fungieren können). Weiterhin wird der Bewertung der Traumafolgen mehr Raum eingeräumt als bisherigen Programmen und die Bearbeitung eines größeren Spektrums verhal-

tenbezogener und kognitiver Vermeidungsstrategien einbezogen.

Ehlers et al. gingen nun in zwei Schritten vor, um die Wirksamkeit ihrer Therapiemodifikation zu testen (Kombination von Phase-II- und Phase-III-Studien). Erst wurde eine fortlaufende Pilotstudie von Einzelfällen (n=20) durchgeführt, um dann nach erfolgreichem Abschluss – die Ergebnisse der Einzelfallstudien waren sehr ermutigend, insbesondere weil die Drop-out-Rate gering war – eine echte Wirksamkeitsstudie mit 28 weiteren Patienten anzuschließen (kognitive Therapie versus Wartegruppe).

Die Ergebnisse waren, wie bei dieser Arbeitsgruppe üblich, ausgesprochen positiv. Sogar anhand der Intention-to-treat-Analysen zeigten sich extrem hohe Effektstärken von deutlich über 2,0; diese Effektstärken übersteigen sogar diejenigen Effekte aus Studien, bei denen nur die Therapiebeender analysiert wurden (d. h. ohne ITT-Prinzip). Zudem führten Ehlers et al. weitere Analysen durch, um Prädiktoren für den Therapieerfolg zu identifizieren. Interessanterweise war die Behandlung über viele Variablen hinweg (z. B. Alter, Geschlecht, Zeit-

raum seit Trauma, Traumaschwere, Komorbidität mit depressiven oder Angststörungen) gleich gut wirksam. Sie war signifikant wirksamer bei niedrigerem im Vergleich zu höherem sozioökonomischem Status, was dem Vorurteil widerspricht, kognitive Verfahren würden nur bei besonders gebildeten Personen funktionieren. Die Veränderung der Bewertungen scheint eine zentrale Rolle gespielt zu haben, denn die Veränderungen in einem Fragebogen zu speziell bei posttraumatischen Belastungsstörungen anzutreffenden Bewertungsmustern korrelierten mit dem Therapieerfolg zu fast 0,6.

Die Arbeitsgruppen um Anke Ehlers, David Clark und Mitarbeitern haben damit erneut demonstriert, wie ein durch fortgesetzte Serien von Studien erarbeitetes, immer spezifischeres Vorgehen außerordentlich deutliche Effekte produzieren kann.

### 30.4.4 Diagnosetechnologien

Wissenschaft und innovative Forschung, die Technologie vorantreibt, ist im Bereich klinisch-psychologischer Interventionen ebenso hilfreich wie in den Ingenieurwissenschaften oder der Nanophysik. Beim »Deutschen Zukunftspreis 2005«, ausgelobt vom Bundespräsidenten (und dotiert mit 250.000 €), befasste sich z. B. ein nominierendes Forscherteam mit einer Technik, mit der man akustische Signale mit einer »Schallkamera« visuell sichtbar machen bzw. darstellen kann. Damit könnten etwa bei der Entwicklung neuer Motoren diese dahingehend »angeschaut« werden, wo genau die Lärmquellen sitzen, was wiederum die Problemanalyse und -lösung sehr erleichtert. Eine solche visuelle Darstellung, die unmittelbar überzeugt und unseren menschlichen Fähigkeiten zur visuellen Musterbildung entgegenkommt, ist der herkömmlichen Datenaufbereitung in der Akustikforschung (präzise, aber unübersichtliche, umfassende lange Tabellen) überlegen.

In der Klinischen Psychologie sind Technologien, die komplexe Signale anschaulich machen, ebenfalls von großer Bedeutung. So ist z. B. die gute Operationalisierung von angenommenen Wirkfaktoren eine notwendige Bedingung für die Erforschung ihrer tatsächlichen Wirkung. Eine umfassende, möglichst spezifisch an den angenommenen Wirkfaktoren ausgerichtete Diagnostik liefert allerdings hoch komplexe Befunde, die auf geeignete Weise – möglichst graphisch – aufbereitet werden müssen, damit ein Behandler daraus unmittelbar Schlüsse für das weitere Vorgehen ableiten kann. Auch die Rückmeldungen über die Effekte der eingesetzten Interventionen sind ausgesprochen hilfreich, um das therapeutische Vorgehen zu verbessern; dies gilt sowohl für die Erforschung und Ent-

wicklung im wissenschaftlichen Bereich als auch für die Qualitätssicherung und -entwicklung in der Versorgungspraxis (► Abschn. 30.2.4). Analog zur Schallkamera können wir hier diagnostische Systeme mit einer optimierten Darstellung von komplexen psychologischen Konstellationen und ihren Veränderungen gut gebrauchen. Ein Beispiel dafür ist die sog. Figurationsanalyse (Grawe & Braun, 1994).

Eine beispielhafte Weiterentwicklung aus dem neurobiologischen Bereich ist die »Neuropattern-Analyse« (Hellhammer, 2005a), einer Methode, die auf Einzelfallebene neurobiologische Schnittstellen und deren Aktivität identifiziert, über welche die Kommunikation zwischen zentralem Nervensystem und Körperorganen bei Stress abläuft. »Neuropattern« eignet sich für induktive und deduktive Untersuchungen und kann als beträchtliche Verbesserung im Bereich der Diagnostik psychosomatischer und somatoformer Störungen betrachtet werden.

## 30.5 Die perfekte therapeutische Intervention

Die Expositionstherapien bei Ängsten zählen zu den relativ gut wissenschaftlich untersuchten und hinsichtlich der Wirksamkeit mit vielen Belegen für viele Störungsbereiche ausgestatteten klinisch-psychologischen Interventionen (► Kap. 24). Doch selbst diese Verfahren – wie bereits am Beispiel der systematischen Desensibilisierung oder zur kognitiven Behandlung der posttraumatischen Belastungsstörung aufgezeigt – schreiten in ihrer wissenschaftlichen Fundierung hinsichtlich der genauen Wirkungsweise noch weiter voran. So sind z. B. folgende interessante Fragen noch nicht abschließend geklärt:

- Was sind nun die entscheidenden Wirkfaktoren (zur Debatte stehen z. B. Ansätze auf Grundlage der Habituation, der kognitiven Neubewertung und der Steigerung der Selbstwirksamkeit; detaillierter ► Kap. 24), und wie wirken sie zusammen?
- Wird das ursprüngliche Problemverhalten (z. B. übersteigerte Angstreaktion, Vermeidung) nach erfolgreicher Therapie gelöscht? Wenn nicht (wofür es Hinweise gibt), warum nicht, bzw.: Kann man etwas dagegen tun?
- Warum sind Expositionstherapien manchmal unwirksam? Gibt es Hintergrundbedingungen, die ggf. erst hergestellt werden müssen?
- Wie ist das Dosis-Wirkungs-Spektrum zu beurteilen? Gibt es hinsichtlich Interventionsdauer und -frequenz (ggf. auf den Einzelfall zugeschnittene) optimale Formate, was die Intensität der Behandlung betrifft?

### Implikationen neurobiologischer Forschung für klinisch-psychologische Interventionen

- Psychologische Störungs- und Therapietheorien werden zunehmend auf den Prüfstand gestellt werden. Sie sind nur so lange gültig, wie sie nicht in Widerspruch zu psychobiologischen Erkenntnissen stehen. Wir haben also mit einer neuen Faktenlage zu tun, für deren bestmögliche Erklärung wiederum neue Theorien der nächsten Generation gefunden werden müssen.
- Auf der Grundlage eines stetig wachsenden neurobiologischen Verständnisses (einschließlich der spezifischen genetischen Mechanismen) ist zu erwarten, dass nicht nur hoch spezifische Pharmakotherapie, sondern auch entsprechende klinisch-psychologische Interventionen ein großes Entwicklungspotenzial haben, bei der Manipulation solcher Mechanismen eine wichtige Rolle zu spielen.
- Insbesondere in Bereichen, in denen die psychotherapeutische Ergebnislage unbefriedigend ist (Non-responder, »chronische Fälle«, seltene komplexe Störungsbilder) würden neue diagnostische Methoden hilfreiche Anstöße zur differenziellen bzw. Kontraindikation geben, falls man einen Misserfolg – z. B. einer Expositionstherapie bei neurobiologisch beschreibbarer Habituationsunfähigkeit – vorhersagen könnte.
- Bereits bewährte psychologische Interventionen könnten systematisch durch neue spezifische unterstützende Therapiebausteine in ihrer Wirkung hinsichtlich der Veränderung neuronalen Geschehens optimiert werden (vgl. z. B. Hofmann et al. 2006).

Derlei offene Fragen haben allerdings gute Aussichten, in absehbarer Zeit nicht nur aufgrund der Weiterentwicklung bereits bestehender Forschungsprogramme, sondern auch aus gänzlich neuen Perspektiven heraus neue Antworten zu erhalten. Es gibt bereits Ansätze, die aufzeigen, wie bestimmte soziale Erfahrungen (z. B. elterliches Pflegeverhalten) über eine Kette von biologischen Ereignissen schließlich eine hochspezifische molekularbiologische Veränderung bewirken, die kausal für eine erhöhte Stressreaktivität verantwortlich ist. Hellhammer (2005b) stellt in einer Würdigung der kanadischen Arbeitsgruppe um Michael Meaney heraus, dass es damit erstmals gelungen ist, die molekularbiologischen Grundlagen aufzudecken, über die soziale Erfahrungen krankheitsrelevant werden.

Der enorme Zuwachs an Wissen über neuronale Korrelate klinisch relevanter Probleme und die zunehmenden Erkenntnisse über neurobiologische Regelkreise und Mechanismen, wie sie z. B. auch bei Grawe (2004) dargestellt sind, haben dabei für die Entwicklung und Beurteilung klinisch-psychologischer Interventionen (sowie auch innovativer diagnostischer Verfahren) mehrere Implikationen.

Wie die Vermittlung von Umwelt- und sozialen Faktoren genau funktioniert, über die etwa die oben genannten biologischen Ereignisketten ausgelöst werden, oder ob und wie wir als Subjekte unser neuronales Geschehen steuern, wird nicht in naher Zukunft abschließend geklärt werden können. Auch die immer differenzierteren neurobiologischen Erkenntnisse bestehen im Gesamtzusammenhang eher aus probabilistischen denn aus kausalen Aussagen, und auch gute derartige Erklärungsmodelle behaupten nicht, unser Erleben und Ver-

halten sei ausschließlich »neurobiologisch determiniert«. Einige Grundfragen (z. B. »Wie kommt die Angst in die Amygdala?«) werden also auch trotz zunehmend größerer Daten- und theoretischer Basis nach wie vor eine große Herausforderung bleiben. Von einer »unifying theory« zu träumen, aus der anhand einer vollständigen Diagnostik die jeweils individuelle perfekte kausale Intervention abgeleitet werden kann, ist unrealistisch, aber dennoch menschlich verständlich (genetisch angelegter optimistischer Bias).

Der Prozess der Entwicklung und Beurteilung therapeutischer Verfahren wird noch auf absehbare Zeit nach dem in diesem Kapitel dargestellten Muster ablaufen: Entwicklung neuer therapeutischer Verfahren anhand von Hypothesen, die möglichst aus der derzeitigen Faktenlage abgeleitet sind – Nachweis von Wirksamkeit und Effektivität – Modifikation des Ansatzes bei neuer Faktenlage – erneute Evaluation etc. Dabei muss stets auch der Aspekt beachtet werden, dass wirksame Interventionen auch in den Versorgungsalltag Eingang finden (vgl. Barlow, 2004). Innovation und Legitimation sollten sich dabei komplementär ergänzen, denn die Wirksamkeit (und Verträglichkeit) neuer Verfahren muss insbesondere dann abgesichert sein, wenn die theoretischen Grundlagen noch nicht auf allen Ebenen vollständig bekannt sind und bei der Entwicklung eher heuristisch vorgegangen wurde.

## Frage

## Prüfen Sie Ihr Wissen

1. Die Forschung zu klinisch-psychologischen Interventionen befasst sich sowohl mit der Beurteilung bzw. Legitimation bereits bestehender Behandlungsprogramme als auch mit deren (Weiter-)Entwicklung. Erläutern Sie diese beiden Aspekte anhand von Beispielen!
  - ▶ Abschn. 30.1.1., 30.1.2., 30.2.5., 30.4.3., 30.5
2. Innerhalb der vier Phasen der Therapieevaluation kommt der Phase III – den Wirksamkeitsstudien im engeren Sinne (»randomized controlled trials«, RCT) – eine besondere Bedeutung zu. Stellen Sie sich vor, Sie hätten alle Mittel zur Verfügung, um ein neuartiges Therapieprogramm zur Behandlung von Drogenabhängigkeit in einer Phase-III-Studie zu prüfen. Erläutern Sie an diesem Beispiel die folgenden Begriffe:
  - Ein- und Ausschlusskriterien
  - Treatmentintegrität
  - Randomisierung
  - Verblindung
  - Primäre und sekundäre Erfolgskriterien
  - Poweranalyse
  - Standardisierung versus Individualisierung
  - »Intention-to-treat« (ITT), »drop-outs«
  - Katamnese, Follow-up
  - CONSORT-Statement
  - »Allegiance«
  - Identifikation von Wirkfaktoren
3. Analog zur »evidence based medicine« wurde in den 1990er Jahren ein Konzept der »empirically supported treatments« für klinisch-psychologische Interventionen entwickelt. Damit sollte – störungsspezifisch – eine Liste »wissenschaftlich anerkannter Verfahren« erstellt werden. Stellen Sie Pro und Contra dieses Ansatzes dar!
  - ▶ Abschn. 30.3.2
4. Eine Methode, die Vielzahl von Einzelergebnissen aus Studien zur Interventionsforschung zu integrieren, ist die Metaanalyse.
  - a) Nennen Sie die Gründe, die Metaanalysen sinnvoll und vorteilhaft erscheinen lassen!
  - b) Nennen Sie typische Probleme, mit denen Metaanalysen zu kämpfen haben, einschließlich möglicher Lösungsansätze!
  - ▶ Abschn. 30.3.1
5. Viele Psychotherapiestudien, die ganz unterschiedliche Verfahren hinsichtlich Wirksamkeit und Effektivität geprüft haben, kamen zu ähnlichen Ergebnissen. Heißt das, dass alle diese Verfahren gleich wirksam sind? Oder dass ihrer Wirksamkeit lediglich allgemeine unspezifische Faktoren zu Grunde liegen? Oder dass Psychotherapie im Grunde ein »Placebo« darstellt?
  - ▶ Abschn. 30.3.1., 30.4.1

## Literatur

### Weiterführende Literatur

Das Wechselspiel von Entwicklung und Beurteilung von Verfahren wird besonders bei Grawe (1995) herausgearbeitet. Bandura (1977) entwirft einen beispielhaften Ansatz, frühere theoretische Ansätze in einer »Theorie zweiter Generation« bzw. einer »unifying theory« zu integrieren. Grawe (1998) legt in seinem Buch »Psychologische Therapie« ebenfalls eine allgemeine Theorie vor, die den Anspruch erhebt, nicht mehr auf Therapieschulen, sondern ausschließlich auf bisheriger empirischer Therapieforschung verschiedenster Couleur sowie auf psychologischer Grundlagenforschung begründet zu sein.

Der Erfolg störungsspezifischer psychologischer Interventionen – gerade auch im Vergleich organmedizinischer Behandlungen – wird von Barlow (2004) besonders gut herausgestellt, weshalb dieser Aufsatz auch berufspolitisch bedeutsam ist. Die Debatte um solche störungsspezifischen Verfahren – also um »empirically supported treatments« (EST) – kann nachgezeichnet werden anhand von Seligman (1995), der ein Plädoyer für Effectiveness-Studien als Ergänzung zu Efficacy-Studien hält, Hahlweg (1995), der die erste Version zu EST der APA-Taskforce ins Deutsche überträgt, sowie Westen, Novotny und Thomson-Brenner (2004), die einen besonders elaborierten – wenn auch ein wenig zu pessimistischen – Überblick zu Methodik und Forschungsstand der Evidenzbasierung psychologischer Verfahren geben. Zum Thema »Metaanalysen« eignet sich als Vertiefung besonders die Übersicht von Rosenthal und Di Matteo (2001). Einen Überblick über das bislang vernachlässigte Feld negativer Outcomes im Zuge von Psychotherapien geben Barlow (2010) sowie Lilienfeld (2007).

Zwei herausragende Therapiestudien, die – obwohl nicht unkompliziert – Fallstricke und Lösungsansätze der Therapieforschung illustrieren, sind die Therapievergleichsstudie zur Panikstörung von Barlow, Gorman, Shear und Woods (2000) und das »National Institute of Mental Health Treatment of Depression Collaborative Research Program«, von dessen umfangreichen Publikationen hier Shea et al. (1992) empfohlen wird.

Da das wichtige »Handbook of Psychotherapy and Behavior Change« als Nachschlagewerk zwar eine zentrale Bedeutung hat, aber hier als Literaturangabe zu umfangreich ist, sei hierfür als Einstieg auf das Zusammenfassungs- und Ausblickskapitel von Bergin und Garfield (2004) verwiesen. Und was die Therapieforschung der Praxis als Entscheidungshilfe im konkreten Einzelfall bieten kann, wird beispielhaft bei Westmeyer (2000) erörtert.

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- 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., Gorman, J.M., Shear, M.K. & Woods, S.W. (2000). Cognitive-behavioral therapy, imipramine, or their combination for panic disorder. A randomized controlled trial. *Journal of the American Medical Association*, 283(19), 2529–2536
- Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M. et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behaviour change consortium. *Health Psychology*, 23, 443–451.
- Bergin, A.E. & Garfield, S.L. (2004). Overview, trends, and further issues. In M.J. Lambert (Eds.), *Bergin & Garfield's Handbook of psychotherapy and behavior change* (5th ed., pp. 805–821). New York: Wiley.
- Borkovec, T.D. & Castonguay, L.G. (1998). What is the scientific meaning of empirically supported therapy? *Journal of Clinical and Consulting Psychology*, 66(1), 136–142.
- Buchkremer, G. & Klingberg, S. (2001). Was ist wissenschaftlich fundierte Psychotherapie?: Zur Diskussion um Leitlinien für die Psychotherapieforschung. *Der Nervenarzt*, 72(1), 20–30.
- Caspar, F. & Jacobi, F. (2004). Psychotherapieforschung. In W. Hiller & E. Leibing (Hrsg.), *Lehrbuch der Psychotherapie* (Band 1, S. 395–410). München: CIP-Medien.
- Chambless, D. & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66(1), 7–18.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S.D. & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine*, 40(2), 211–223.
- Cukrowicz, K.C., White, B.A., Reitzel, L.R., Burns, A.B., Driscoll, K.A., Kemper, T.S. & Joiner, T. E. (2005). Improved treatment outcome associated with the shift to empirically supported treatments in a graduate training clinic. *Professional Psychology: Research and Practice*, 36(3), 330–337.
- Ehlers, A., Clark, D.M., Hackmann, A., McManus, F. & Fennell, M. (2005). Cognitive therapy for post-traumatic stress disorder: Development and evaluation. *Behaviour Research and Therapy*, 43, 413–431.
- Eysenck, H.J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16, 319–324. [Nachdruck in: *Journal of Consulting and Clinical Psychology*, 60(5), 659–663 (October 1992)].
- Flückinger, C., Regli, D., Zwahlen, D., Hostettler, S. & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. Ein Instrument zur Erfassung von Therapieprozessen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 39(2), 71–79.
- Fonagy, P. & Roth, A. (2004). What works for whom?: A critical review of psychotherapy research (2. Aufl.). New York: Guilford.
- Frank, J.D. (1961). *Persuasion and healing*. Baltimore: The Johns Hopkins University Press.
- Franklin, M.E., Abramowitz, J.S., Kozak, M.J., Levitt, J.T. & Foa, E.B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology*, 68 (4), 594–602.

- Grawe, K. (1995). Grundriss einer Allgemeinen Psychotherapie. *Psychotherapeut*, 40(3), 130–145.
- Grawe, K. (1997). Research informed psychotherapy. *Psychotherapy Research*, 7(1), 1–19.
- Grawe, K. (1998). *Psychologische Therapie*. Göttingen: Hogrefe.
- Grawe, K. (2004). *Neuropsychotherapie*. Göttingen: Hogrefe.
- Grawe, K. & Braun, U. (1994). Qualitätskontrolle in der Psychotherapiepraxis. *Zeitschrift für Klinische Psychologie*, 23(4), 242–267.
- Grawe, K., Donati, R. & Bernauer, F. (1994). *Psychotherapie im Wandel: Von der Konfession zur Profession*. Göttingen: Hogrefe.
- Hahlweg, K. (1995). Zur Förderung und Verbreitung psychologischer Verfahren. Ein APA-Bericht. *Zeitschrift für Klinische Psychologie*, 24(4), 275–284.
- Hahlweg, K., Fiegenbaum, W., Frank, M., Schröder, B., & von Witzleben, I. (2001). Short- and longterm effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology*, 69, 375–382.
- Hartmann, A. & Herzog, T. (1995). Varianten der Effektstärkenberechnung in Meta-Analysen: Kommt es zu variablen Ergebnissen? *Zeitschrift für Klinische Psychologie*, 24(4), 337–343.
- Hellhammer, D. (2005a). Neuropattern – ein innovatives psychobiologisches Verfahren zur Prävention, Diagnostik und Behandlung von stressbezogenen Gesundheitsstörungen. *Report Psychologie*, 30(7/8), 312–316.
- Hellhammer, D. (2005b). Interview mit Michael Meaney: Wie die Zuwendung der Eltern die Stressvulnerabilität beeinflusst: Molekularbiologische Grundlagen sozialer Erfahrung. *Verhaltenstherapie*, 15, 110–112.
- Hiller, W., Fichter, M.M., Rief, W. (2003) A controlled treatment study of somatoform disorders including analysis of health care utilization and cost-effectiveness. *Journal of Psychosomatic Research*, 54, 369–380.
- Hiller, W., Bleichhardt, G. & Schindler, A. (2009). Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 57, 7–21.
- Höfler, M., Gloster, A.T. & Hoyer, J. (2010). Causal effects in psychotherapy: Counterfactuals counteract overgeneralization. *Psychotherapy Research*.
- Hofmann, S.G., Meuret, A.E., Smits, J.A., Simon, N.M., Pollack, M.H., Eisenmenger, K. et al. (2006). Augmentation of exposure therapy with D-Cycloserine for social anxiety disorder. *Archives of General Psychiatry*, 63, 298–304.
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Humans (ICH). (1998). *Guideline E9: Statistical principles for clinical trials*. Verfügbar unter: <http://www.ich.org> [07.04.2011].
- Jacobson, N.S., Roberts, L.J., Berns, S.B., McGlinchey, J.B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300–307.
- Kendall, P.C. (1998). Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, 66(1), 3–6.
- Kiesler, D.J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, 65, 110–136.
- Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J. & Johnson, B.T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the food and drug administration. *PLoS Medicine*, 5, 260–268.
- Kopta, S.M., Howard, K.I., Lowry, J.L. & Beutler, L.E. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62(5), 1009–1016.
- Lambert, M.J. (Ed.) (2004). *Bergin & Garfield's Handbook of psychotherapy and behavior change* (5th ed.). New York: Wiley.
- Lambert, M.J. (2005). Emerging methods for providing clinicians with timely feedback on treatment effectiveness: An introduction. *Journal of Clinical Psychology/In Session*, 61(2), 141–144.
- Leichsenring, F. & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy A meta-analysis. *Journal of the American Medical Association*, 300, 1551–1565.
- Liberati, A., Altman, D.G., Tetzlaff, J. et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10), e1–34
- Lilienfeld, S.O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53–70.
- Lincoln, T.M. (2010). Letter to the Editor: A comment on Lynch et al. (2009). *Psychological Medicine*, 40(5), 877–880.
- Lincoln, T.M., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B. & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy*, 41, 1251–1269.
- 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., Martinovich, Z. & Howard, K.I. (2001). Vorhersage individueller Psychotherapieverläufe. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 30, 104–113.
- Lynch, D., Laws, K.R. & McKenna, P.J. (2009). Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. *Psychological Medicine*, 40(1), 9–24.
- Malik, M.L., Beutler, L.E., Alimohamed, S., Gallagher-Thompson, D. & Thompson, L. (2003). Are all cognitive therapies alike? A comparison of cognitive and noncognitive therapy process and implications for the application of empirically supported treatments. *Journal of Consulting and Clinical Psychology*, 71(1), 150–158.
- Märtens, M. & Petzold, H. (Hrsg.). (2002). *Therapieschäden*. Mainz: Grünewald.
- Margraf, J., Ehlers, A., Taylor, C.B., Roth, W.T., Clark, D.B., Sheikh, J. & Agras, W.S. (1991). How »blind« are double-blind studies? *Journal of Consulting and Clinical Psychology*, 59, 184–187.
- Margraf, J. & Lieb, R. (1995). Was ist Verhaltenstherapie? Versuch einer zukunftsorientierten Neucharakterisierung. Editorial. *Zeitschrift für Klinische Psychologie*, 24, 1–7.
- Moher, D., Schulz, K.F., Altman, D.G. für die CONSORT-Gruppe (2004). Das CONSORT Statement: Überarbeitete Empfehlungen zur Qualitätsverbesserungen von Reports randomisierter Studien im Paralleldesign. *Deutsche Medizinische Wochenschrift*, 129, 16–20.
- Müller-Oerlinghausen, B. & Linden, M. (1981). Rationalität der Indikation zu psychopharmakologischer Behandlung. In: U. Baumann (Hrsg), *Indikation zur Psychotherapie* (S. 210–220). München: Urban & Schwarzenberg.
- Nathan, P. E., & Gorman, J. M. (Eds). (2002). *A guide to treatments that work* (2nd ed.). New York: Oxford University Press.
- Norcross, J. (Ed.). (2003). *Psychotherapy relationships that work*. New York: Oxford University Press.
- Perst A. & Baumann, U. (1999). Einzelfallstudien in klinisch-psychologischen, psychotherapeutischen Fachzeitschriften. *Zeitschrift für Klinische Psychologie*, 28, 205–213.
- Plath, I. (1998). Die Untersuchung von Grawe, Donati & Bernauer (1994) aus forschungsintegrativer Sicht. Eine methodenkritische Analyse. *Report Psychologie*, 23(9), 730–749.

- Rief, W. & Hofmann, S.G. (2009). Dies Psychoanalyse soll gerettet werden. Mit allen Mitteln? *Nervenarzt*, 80(5), 593–597.
- Rief, W., Nestoriuc, Y., Weiss, S., Welzel, E., Barsky, A.J. & Hofmann, S.G. (2009). Meta-analysis of the placebo response in antidepressant trials. *Journal of Affective Disorders*, 118, 1–8.
- Rosen, G.M. & Davison, G.C. (2003). Psychology should list empirically supported principles of change (ESPs) and not trademarked therapies or other treatment packages. *Behavior Modification*, 27(3), 300–312.
- Rosenthal, R. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169.
- Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist*, 45, 775–777.
- Rosenthal, R. & Di Matteo, M.R. (2001). Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52(1), 59–82.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71–72.
- Schulte, D. (1993). Lohnt sich eine Verhaltensanalyse? *Verhaltenstherapie*, 3, 5–13.
- Seligman, M.E.P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, 50(12), 965–974.
- 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.
- Shadish, W.R., Matt, G.E., Navarro, A.M. & Siegle, G. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, 65(3), 355–365.
- Shea, M.T., Elkin, I., Imber, S.D., Sotsky, S., Watkins, J., Collins, J. et al. (1992). Course of depressive symptoms over follow-up – Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, 49(10), 782–787.
- Smith, M.L., Glass, G.V. & Miller, T.I. (1980). Meta-analysis of psychotherapy. *American Psychologist*, 41, 165–180.
- Southworth, S. & Kirsch, I. (1988). The role of expectancy in exposure-generated fear reduction in agoraphobia. *Behaviour Research and Therapy*, 26, 113–120.
- Stewart-Williams, S. (2004). The placebo puzzle: Putting together the pieces. *Health Psychology*, 23(2), 198–206.
- Wade, W., Treat, T. & Stuart, G. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231–239.
- Westen, D., Novotny, C.M. & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, Findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 130(4), 631–663.
- Westmeyer, H. (2000). Wissenschaftstheoretische Aspekte. In: J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie* (Band 1, S. 31–48). Berlin: Springer.
- Wolpe, J. (1958). *Psychotherapy by reciprocal inhibition*. Stanford: Stanford University Press.
- Zielke, M. (1993). *Wirksamkeit stationärer Verhaltenstherapie*. Weinheim: Beltz, Psychologie Verlags Union.