

Evidenzbasiertes Arbeiten in der Physio- und Ergotherapie

► Zusammenfassungen/Abstracts

Das Wichtigste zu jedem einzelnen Kapitel in Kürze.
Zum Nachschlagen, auffrischen oder einfach als Merktzettel.

► Zum Weiterlesen

Wen es jetzt richtig gepackt hat, der kann noch mehr zu evidenz-basiertem Arbeiten in den folgenden Kapiteln erfahren.
Die praktischen Zusammenfassungen gibt es hier natürlich auch.

► Übungsblätter

Sämtliche Übungsblätter und das Excelfile „Berechnung der EBP-Größen“ finden Sie unter <http://extras.springer.com> (978-3-642-40635-5).



Diese Publikation ist ein ergänzendes, frei zugängliches Dokument zum Buch:

Mangold S (2013) Evidenzbasiertes Arbeiten in der Physio- und Ergotherapie. Reflektiert - systematisch - wissenschaftlich fundiert, 2. Aufl. Springer, Berlin Heidelberg

Bitte zitieren Sie diese Publikation folgendermaßen:

Mangold S (2013) <http://extras.springer.com>. Download
zu: Evidenzbasiertes Arbeiten in der Physio- und Ergotherapie. Reflektiert - systematisch - wissenschaftlich fundiert,
2. Aufl. Springer, Berlin Heidelberg

Zusammenfassungen/Abstracts

Buchkapitel

Kapitel 1 – Methoden der evidenzbasierten Arbeit Zusammenfassung

Die therapeutischen Ansätze basieren auch heute noch z.T. ausschließlich auf theoretischen Überlegungen, Beobachtungen sowie auf Versuch und Irrtum. Grundsätzlich ist es sinnvoll, anhand solcher theoretischen Überlegungen etc. Therapieansätze zu entwickeln, jedoch muss die wissenschaftliche Überprüfung folgen, welche wissenschaftliche Evidenz generiert.

Die Bedeutung des Begriffs Evidenz umfasst eine Spannweite von der Augenscheinlichkeit bis hin zum wissenschaftlichen Beweis, wenn auch Letzterer stets mit einer Unsicherheit behaftet ist. Dementsprechend gibt es verschiedene Methoden, Evidenz mit unterschiedlicher Beweiskraft zu sammeln. Dazu gehören, in der Reihenfolge steigenden Evidenzniveaus aufgeführt

- die reflektierte Praxis,
- die systematischen Beobachtungen am Patienten,
- die EBP.

Auch die angewandte Forschung bietet eine Möglichkeit, selbst Evidenz zu sammeln. Sie bedarf jedoch zusätzlicher Stellenprozente für wissenschaftlich geschulte Therapeuten oder Wissenschaftler.

Kapitel 2 – Reflektierte Praxis Zusammenfassung

Die reflektierte Praxis entwickelte Donald A. Schon in den 1980er-Jahren für die Pädagogik. Sie eignet sich unter entsprechenden Anpassungen auch für die therapeutische Arbeit und dient dazu, unbefriedigende und unpassende Vorgehensweisen aufzudecken sowie individuelle psychologische Faktoren des Patienten und des Therapeuten im Kontext der Therapie zu analysieren. Ziel ist, die Therapiesituation besser zu verstehen und das Miteinander zwischen Patient und Therapeut zu erleichtern.

Zur reflektierten Praxis gehören mehrere Schritte, d.h. die Wahrnehmung des Geschehens in der Therapiesituation, die Reflexion während und nach der Therapie, die Umsetzung der Evidenz in die Praxis, die Überprüfung der Wirksamkeit sowie Überlegungen, ob sich daraus Schlüsse für zukünftige Patienten ziehen lassen. Das Konzept beinhaltet eine Reihe von Fragen, die dem Therapeuten helfen, diese Schritte durchzuführen.

Die Evidenz ist eher subjektiver als objektiver Natur. Sie lässt sich nicht verallgemeinern, bietet aber eine wertvolle Hilfe für die Therapie des betreffenden Patienten.

Kapitel 3 – Systematische Beobachtungen am Patienten Zusammenfassung

Systematische Beobachtungen am Patienten dienen dazu, sich einen Eindruck über den Erfolg oder Misserfolg einer Intervention beim eigenen Patienten zu verschaffen oder aus mehreren Interventionen eine geeignete für ihn und mit ihm auszuwählen. Wichtig ist, systematisch dabei vorzugehen, sich im möglichen Rahmen solche Gedanken zu machen, welche auch wissenschaftlichen Experimenten zugrunde liegen. Wichtige Punkte sind dabei, sich zu überlegen:

- welche Größe beim Patienten zu erfassen ist, damit der Therapeut den Therapieerfolg überprüfen kann (abhängige Variable),

- wodurch eine Veränderung erreicht werden soll (unabhängige Variable),
- in welchem Zeitraum sich überhaupt eine sichtbare Veränderung ereignen kann,
- welche Größen die Veränderung noch beeinflussen könnten (Störvariablen),
- wie sich Störvariablen bei Testwiederholungen möglichst konstant halten oder ausschalten lassen, um klarere Resultate zu erzielen.

Kapitel 4 – Herkunft und Konzept der evidenzbasierten Praxis Zusammenfassung

Die EBP hat ihre Wurzeln in der EBM. Diese fand ihrerseits ihren Ursprung in der Mitte des 19. Jahrhunderts mit dem Ziel, medizinische Konzepte aufgrund von Erfahrungen in Konzepte mit objektiver Beweislage umzuwandeln.

Die EBM und EBP verwenden wissenschaftliche Studien. In der Medizin gibt es seit einiger Zeit viel mehr Studien, als der Praktizierende lesen kann. Zusammenfassende Studien auf wissenschaftlichem Niveau, z.B. systematische Übersichtsartikel, Meta-Analysen und Leitlinien, geben einen wertvollen Überblick über die Ergebnisse wichtiger Einzelstudien. Dadurch kann der Praktizierende Zeit sparen, denn er braucht die betreffenden Einzelstudien nicht zu lesen. In der Therapie gibt es bisher bedeutend weniger Studien, aber ihre Anzahl steigt exponentiell an.

Die EBP stützt sich zwar auf die wissenschaftlichen Erkenntnisse (externe Evidenz), verzichtet aber keinesfalls auf den praktischen Erfahrungsschatz des Therapeuten (interne Evidenz). Beide zusammen führen zu besseren Entscheidungen hinsichtlich der Behandlung des Patienten.

Die praktische Durchführung der EBP beginnt mit der Problemidentifizierung und Formulierung der Fragestellung. Darauf folgt eine Literaturrecherche und kritische Literaturarbeit. Bei der Einschätzung der Anwendbarkeit der Evidenz berücksichtigt der Therapeut auch seine eigenen Erfahrungen und die Wünsche des Patienten. Schließlich wendet der Therapeut die Erfolg versprechende Maßnahme an und überprüft, ob sie das gewünschte Ergebnis beim Patienten erzielt.

Kapitel 5 – Problemidentifizierung und Formulierung der Fragestellung Zusammenfassung

Der 1. Schritt der EBP ist die Problemidentifizierung. Sie dient dazu, einen Wissensbedarf bei der Behandlung eines Patienten aufzudecken. Sie erfolgt durch den Patienten selbst, der z.B. Zweifel an der angewendeten Methode anmelden kann, oder durch den aufmerksamen Therapeuten, welcher beispielsweise mithilfe der reflektierten Praxis Problemstellungen erkennt.

Die Problemidentifizierung ist noch auf zu ungenauer Ebene, um sich damit gezielt auf die Literatursuche begeben zu können. Dazu braucht es präzise Fragen. Mithilfe des Akronym PICO kann man sich gut merken, welche Elemente in die Fragen gehören. P steht für Patient oder Problem, I bedeutet Intervention, C Comparison (Vergleichsbehandlung) und O Outcome (Zielgröße). Wenn nur eine Intervention im Zentrum des Interesses steht, darf die Vergleichsbehandlung in der Fragestellung fehlen. Die PICO-Elemente sind auf

alle Themenbereiche anwendbar, z.B. bei Fragen zur Wirksamkeit therapeutischer Interventionen, Prognose oder Ätiologie.

Kapitel 6 – Literaturrecherche

Zusammenfassung

Die Verwendung elektronischer Datenbanken und Suchmaschinen im Internet bilden die einfachste und effizienteste Art, Studien zu suchen. Dazu zählen z.B. PubMed, PEDro und OT seeker. Um die Suche auf die eigene Fragestellung zu fokussieren und ungeeignete Artikel auszufiltern, eignen sich MeSH, Operatoren, tags oder z.B. das Fenster »Limits« bei PubMed.

Neben dem Internet existieren weitere Möglichkeiten, Studien zu suchen. Darunter lassen sich zwei besonders empfehlen: Experten auf dem betreffenden Gebiet nach Literatur zu fragen und Literaturverzeichnisse passender Studien durchzusehen.

Manche Volltexte der Studien können kostenlos auch ohne Lizenzen direkt aus dem Internet ausgedruckt werden. Andere lassen sich im Internet kostenpflichtig bestellen. Öffentliche Bibliotheken bieten weitere Möglichkeiten, sich Zugang zu den Artikeln zu verschaffen.

Kapitel 7 – Einführung in die Beurteilung wissenschaftlicher Studien

Zusammenfassung

Studien unterscheiden sich in ihrer wissenschaftlichen Güte. Neben dem Studiendesign gibt der Aufbau eines Artikels erste Hinweise darüber. Entspricht der Artikel nicht der typischen Struktur wissenschaftlicher Publikationen (Einleitung, Methoden, Ergebnisse, Diskussion), so lässt er sich für die EBP nicht auswerten.

Gütekriterien, in der EBP als Leitfragen formuliert, geben genauere Auskunft über das wissenschaftliche Niveau der Studien. Dazu gehören z.B. die Verblindung der Versuchsteilnehmer und der behandelnden Therapeuten, die Randomisierung und die Intention-to-Treat-Analyse. Doch gilt es nicht nur, die wissenschaftliche Glaubwürdigkeit der Studie einzuschätzen, sondern auch die Relevanz der Ergebnisse zu beurteilen. Dazu sind besonders Größen wie der minimale klinisch wichtige Unterschied, die Effektgröße und die Number Needed to Treat zu beachten. Zuletzt ist zu überlegen, ob sich die Evidenz in der Praxis anwenden lässt, d.h. ob sich die Ergebnisse auf den eigenen Patienten übertragen lassen und ob es möglich ist, die Intervention in der Praxis anzuwenden.

Kapitel 8 – Modelle als Grundlage für die Befunderhebung

Zusammenfassung

Eine gute Grundlage für die Befunderhebung bieten Modelle, aus denen Therapeuten und Therapeutinnen Befundbögen ableiten können. Der Befund soll dazu dienen, passende Therapieziele zu formulieren, sinnvolle Messungen zu planen und geeignete Mittel und Methoden zu wählen. Um diesen Zielen gerecht zu werden, muss das Modell valide, praxisrelevant und anwendbar sein. Diese 3 Eigenschaften überprüft der Therapeut mithilfe vorgegebener Leitfragen.

- Bei der Validität wird beispielsweise nach den theoretischen Grundlagen, auf welchen das Modell basiert, und seiner inneren Konsistenz gefragt.
- Bei der Überprüfung der Relevanz achtet der Therapeut z.B. darauf, ob das Modell bei der Aufdeckung sinnvoller Ziele hilft, wissenschaftlich begründete Vorschläge für Therapieansätze oder Tests bietet und Vorhersagen trifft.
- Zur Beurteilung der Anwendbarkeit beantwortet der Therapeut Leitfragen dazu, ob der Bezugsrahmen zur Problemstellung seines Patienten passt, ob dieser der Zielgruppe des Modells entspricht und ob das Modell für die Praxis nicht zu komplex ist.

Manche Modelle sind nicht nur für die eigene Berufsgruppe, sondern für den interdisziplinären Gebrauch bestimmt. In diesem Fall ist bei der Wahl des Modells zusätzlich zu den oben aufgeführten

Punkten z.B. darauf zu achten, dass es für alle verständlich ist und alle beteiligten Berufsgruppen davon profitieren.

Kapitel 9 – Kritische Bewertung ätiologischer Studien

Zusammenfassung

Ätiologische Studien dienen dazu, Krankheitsursachen aufzudecken. Dafür eignen sich bei den Primärstudien besonders RCTs und Kohortenstudien. Sie vergleichen die Ergebnisse einer Gruppe, die einem potentiell belastenden Faktor ausgesetzt ist (Expositionsgruppe) mit denjenigen einer Kontrollgruppe.

Da sich die Studien hinsichtlich ihrer wissenschaftlichen Güte unterscheiden, muss der Therapeut zunächst deren Glaubwürdigkeit (Validität) einschätzen. Das erfolgt mithilfe bestimmter Leitfragen, die u.a. nach der Gleichartigkeit der Bedingungen in der Expositions- und Kontrollgruppe, nach der Verblindung und nach dem logischen Zusammenhang zwischen Ursache und Wirkung fragen.

Anschließend muss der Therapeut herausfinden, wie deutlich die Krankheit durch die Exposition zustande kommt. Dazu gibt es mehrere Kenngrößen wie das RR, die RO, RRI, ARI und die NNT_H, von denen letztere Größe die anschaulichste ist.

Zum Schluss muss der Therapeut ermitteln, inwiefern sich die Evidenz bei dem zu behandelnden Patienten in die Praxis umsetzen lässt. Dazu schätzt er ab, wie sehr sich der Patient von denjenigen in der Studie unterscheidet und passt ggf. die NNT_H für seinen Patienten in Form der NNT_{H-ind} an. Zudem hinterfragt er, welche praktische Konsequenzen es für den Patienten hätte, würde er die Exposition vermindern oder vermeiden und ob der Patient dazu bereit ist.

Kapitel 10 – Kritische Bewertung prognostischer Studien

Zusammenfassung

Studien zur Prognose untersuchen die Entwicklung einer Krankheit in einer definierten Zeitspanne hinsichtlich bestimmter Zielgrößen, z.B. Schmerzen, Grad der Arbeitsfähigkeit, Heilung oder Tod. Die typische und wissenschaftlich anerkannte Form der Primärstudien ist die Kohortenstudie. Manche Studien untersuchen die Prognose für die gesamte Patientengruppe, manche dagegen für Untergruppen, indem sie die Patientengruppe hinsichtlich prognostischer Faktoren aufteilen und getrennt auswerten. Untergruppenanalysen sind wissenschaftlich weniger robust, ermöglichen aber dafür besser angepasste Prognosen.

Da sich die wissenschaftlichen Studien hinsichtlich ihrer Güte unterscheiden, muss der Therapeut zunächst deren Glaubwürdigkeit (Validität) einschätzen. Dazu beantwortet er bestimmte Leitfragen, die u.a. nach einem repräsentativen Patientenkollektiv, nach der Verblindung und nach den Korrekturen bezüglich wichtiger prognostischer Faktoren fragen.

Als Nächstes muss der Therapeut der Studie entnehmen oder berechnen, wie hoch die Wahrscheinlichkeit ist, mit der die Patienten ein bestimmtes Ergebnis in einer spezifischen Zeitspanne erreichen. Zum Schluss muss der Therapeut überlegen, inwiefern sich die Evidenz auf den zu behandelnden Patienten übertragen lässt. Dazu schätzt er die Ähnlichkeit zwischen dem Patientenkollektiv der Studie und dem eigenen Patienten ab. Ferner sind die Vorstellungen des Patienten zu berücksichtigen.

Kapitel 11 – Testarten und deren Gütekriterien

Zusammenfassung

Tests dienen dazu, eine Diagnose zu stellen und den Schweregrad einer Erkrankung oder Behinderung sowie deren Symptome zu erfassen. Tests bzw. Erfassungs- oder Messinstrumente müssen wissenschaftlichen Anforderungen genügen, damit sich der Patient und der Therapeut auf die Ergebnisse verlassen können.

Tests teilen sich in 2 Kategorien auf. Solche der 1. Kategorie sammeln dichotome Daten und beschränken sich somit auf 2 Merkmalsaus-

prägungen, z. B. selbständig/unselbständig. Tests der 2. Kategorie sammeln Daten mit mehr als 2 Merkmalsausprägungen. Ein Beispiel dafür ist eine Schmerzskala von 1–10 Punkten.

Die Objektivität, Reliabilität, Sensitivität, Validität und Praktikabilität sind Gütemerkmale beider Testkategorien, wenn sich auch die Definitionen und Analysemethoden größtenteils unterscheiden. Die Spezifität betrifft nur dichotome Tests.

Kapitel 12 – Tests mit dichotomen Merkmalsausprägungen: Vergleich des Tests mit einem Referenzstandard

Zusammenfassung

Dichotome Tests zeichnen sich dadurch aus, dass sie nur 2 Merkmalsausprägungen unterscheiden, z. B. selbständig/unselbständig oder krank/nicht krank.

Die meisten Tests zeigen nicht bei allen Patienten das richtige Ergebnis an. Wie hoch ihre Trefferquoten und damit ihre Güte sind, decken Kennwerte wie die Sensitivität und Spezifität auf. Die Wissenschaftler überprüfen die Güte eines Tests, indem sie dessen Resultate mit denjenigen eines bewährten Referenzstandards vergleichen.

Um die Eignung eines Tests für die Praxis zu beurteilen, muss der Therapeut wissenschaftliche Artikel, welche sich mit der Überprüfung der wissenschaftlichen Güte des Tests beschäftigen, kritisch lesen und analysieren. Zunächst beurteilt er mithilfe vorgegebener Leitfragen, ob eine Studie glaubwürdig (valide) ist, z. B. ob das Spektrum der Studienteilnehmer repräsentativ für die Patienten in der Praxis ist und ob die Ergebnisse des zu validierenden Tests interpretiert wurden, ohne die Ergebnisse des Referenzstandards zu kennen.

Danach bestimmt der Therapeut, ob die Ergebnisse bedeutsam sind. Ziel der Testanwendung in der Praxis ist, die definitive Krankheitswahrscheinlichkeit (Nachtestwahrscheinlichkeit) für den eigenen Patienten zu ermitteln. Wie präzise deren Wert ist, hängt aber nicht nur von der Güte des Tests, sondern auch von der Genauigkeit ab, mit welcher der Therapeut oder Arzt die Vortestwahrscheinlichkeit des Patienten schätzt.

Abschließend beurteilt der Therapeut mithilfe vorgegebener Leitfragen die Anwendbarkeit des Tests in der Praxis. Dazu überlegt er z. B., inwiefern sich die Evidenz auf den zu behandelnden Patienten übertragen lässt und ob die Kenntnis der Nachtestwahrscheinlichkeit dem Patienten nützt.

Kapitel 13 – Tests mit dichotomen Merkmalsausprägungen: Überprüfung der Objektivität

Zusammenfassung

Die Objektivität, auch Inter-rater-Reliabilität genannt, bezeichnet die Unabhängigkeit der Testresultate vom Untersucher. Wissenschaftler überprüfen die Objektivität, indem sie mindestens 2 Untersucher den Test an denselben Patienten durchführen lassen und überprüfen, wie sehr die Ergebnisse übereinstimmen.

Ein wissenschaftlicher Artikel, welcher die Objektivität eines Tests untersucht, muss glaubwürdig sein. Der Therapeut überprüft die Validität der Studie mithilfe von Leitfragen, z. B. ob das Spektrum der Studienteilnehmer repräsentativ für die Patienten in der Praxis ist und ob jeder Testanwender zu seinen Ergebnissen gelangte, ohne die Resultate der anderen Testanwender zu kennen.

Danach analysiert der Therapeut, ob die Ergebnisse bedeutsam sind. Dazu beachtet er 2 Maße: Die Übereinstimmungsrate, welche sich aus den Daten der Vierfeldertafel berechnet, und den Übereinstimmungskoeffizienten mit zugehöriger statistischer Irrtumswahrscheinlichkeit.

Schließlich beurteilt der Therapeut die Anwendbarkeit des Tests in der Praxis. Dazu beantwortet er Leitfragen, z. B. ob die Art der Versorgung in der Studie mit der eigenen genügend übereinstimmt,

ob relevante Eigenschaften des eigenen Patienten denjenigen der Studienpatienten ausreichend ähneln und ob sich der Test in der Praxis genauso wie in der Studie durchführen lässt.

Kapitel 14 – Tests mit mehr als 2 Merkmalsausprägungen: Überprüfung der Reliabilität

Zusammenfassung

Die Reliabilität ist ein wichtiges Gütemerkmal für Tests mit mehr als 2 Merkmalsausprägungen. Sie beschreibt das Ausmaß, mit welchem ein Test konsistente Daten produziert und sie auch reproduziert, sofern sich bei der Person hinsichtlich der untersuchten Zielgröße nichts geändert hat.

Eine wichtige Art, die Reliabilität zu überprüfen, ist die Intra-rater-Reliabilität. Sie ist das Ausmaß, mit welchem ein Therapeut bei einer Messwiederholung die Ergebnisse der 1. Messung reproduzieren kann.

Ein wissenschaftlicher Artikel, welcher die Intra-rater-Reliabilität eines Tests untersucht, muss glaubwürdig sein. Der Therapeut überprüft deshalb als Erstes die Validität der Studie. Dazu beantwortet er Leitfragen, z. B. ob der zeitliche Abstand zwischen den Messzeitpunkten angemessen war und ob Personen untersucht wurden, welche sich in der untersuchten Zielgröße genügend unterscheiden.

Als Zweites bewertet der Therapeut die Bedeutsamkeit der Ergebnisse, und zwar anhand des Korrelationskoeffizienten, welcher die Enge des Zusammenhangs zwischen den Messreihen angibt, und der zugehörigen Irrtumswahrscheinlichkeit. Je nach Art des Korrelationskoeffizienten kann es zudem notwendig sein, das Ausmaß herauszufinden, mit dem die Werte der Wertepaare übereinstimmen. Dazu sucht er nach Angaben wie die Übereinstimmungsrate oder den Wert der Wiederholbarkeit. Vielleicht gibt es in dem Artikel auch Tabellen oder Grafiken, die Auskunft über die Übereinstimmung geben.

Als Drittes beurteilt der Therapeut die Anwendbarkeit des Tests. Dazu beantwortet er Leitfragen, beispielsweise ob der Patient den Test durchführen und ob das Testergebnis dem Patienten nützen würde.

Eine zweite Art, die Reliabilität zu untersuchen, ist die interne Konsistenz. Sie drückt aus, wie sehr Testteile miteinander korrelieren. Der Therapeut kann sie analog zur Intra-rater-Reliabilität beurteilen.

Kapitel 15 – Tests mit mehr als 2 Merkmalsausprägungen: Überprüfung der Objektivität

Zusammenfassung

Die Objektivität, auch Inter-rater-Reliabilität genannt, gibt das Ausmaß der Zuverlässigkeit an, mit welchem verschiedene Testanwender zu übereinstimmenden Ergebnissen gelangen. Zur Überprüfung der Objektivität wenden mindestens 2 Therapeuten den Test an denselben Patienten an. Deren Werte werden anschließend statistisch miteinander verglichen.

Ein wissenschaftlicher Artikel, welcher die Objektivität eines Tests untersucht, muss glaubwürdig sein. Der Therapeut überprüft die Validität der Studie mit Hilfe von Leitfragen, z. B. ob der zeitliche Abstand zwischen den Tests der verschiedenen Testanwender angemessen war und ob jeder Testanwender zu seinen Ergebnissen gelangte, ohne die Resultate des anderen zu kennen.

Danach beurteilt der Therapeut, ob die Ergebnisse praxisrelevant sind. Dazu dienen ihm der Korrelationskoeffizient, der die Enge des Zusammenhangs zwischen den Messreihen angibt sowie die zugehörige Irrtumswahrscheinlichkeit. Je nach Art des Korrelationskoeffizienten kann es zudem notwendig sein, das Ausmaß herauszufinden, mit dem die Werte der Wertepaare übereinstimmen. Dazu sucht er im Artikel nach passenden Informationen in Tabellen oder Graphiken bzw. nach Angaben wie die Übereinstimmungsrate oder den Wert der Wiederholbarkeit.

Als Letztes beurteilt der Therapeut die Anwendbarkeit des Tests. Dazu beantwortet er Leitfragen, beispielsweise, ob der Test zugänglich und bezahlbar ist und ob die Evidenz einen Einfluss auf das weitere klinisch-therapeutische Vorgehen hat.

Kapitel 16 – Tests mit mehr als

2 Merkmalsausprägungen:

Überprüfung der Validität

Zusammenfassung

Die Validität eines Tests gibt an, wie gut der Test in der Lage ist, das zu messen, was er zu messen vorgibt, ob er angemessen ist, zur Problemstellung passt, der Realität entspricht und das richtige Maß angibt.

Zu den Hauptarten der Validität zählen die Inhalts-, Augenschein-, Kriteriums- und Konstruktvalidität. Die Wissenschaftler prüfen sie mithilfe verschiedener Verfahren: die Augenscheinvalidität im einfachsten Fall durch die subjektive Einschätzung einer Person, die Inhaltsvalidität z.B. mithilfe einer Delphi-Prozedur, die Kriteriumsvalidität anhand eines Referenzstandards und die Konstruktvalidität anhand bereits bekannter Konstrukte bzw. derer gesicherten Beziehungen.

Ein wissenschaftlicher Artikel, welcher die Validität eines Tests untersucht, muss für die Praxis glaubwürdig und seine Ergebnisse praxisrelevant sein. In welchem Maß diese Anforderungen erfüllt sind, untersucht der Therapeut anhand von Leitfragen, deren Inhalt von der Art der Validität abhängt. Über die Relevanz geben z.T. auch statistische Größen (statistische Signifikanz und Korrelationskoeffizient) Auskunft. Zudem beurteilt der Therapeut die Anwendbarkeit des Tests. Dazu beantwortet er Leitfragen, beispielsweise, ob der Test zugänglich und bezahlbar ist und ob die Evidenz einen Einfluss auf das weitere klinisch-therapeutische Vorgehen hat.

Kapitel 17 – Tests mit mehr als

2 Merkmalsausprägungen:

Überprüfung der Sensitivität

Zusammenfassung

Die Sensitivität bezeichnet die Fähigkeit eines Erfassungsinstrumentes, Veränderungen der Zielgröße bei der Patientengruppe, für welche es bestimmt ist, innerhalb eines angemessenen Zeitraumes nachzuweisen.

Wissenschaftler ermitteln die Sensitivität durch (mindestens) 2 Messungen pro Patient, zwischen welchen eine genügend große Zeitspanne liegen muss, damit sich eine Veränderung ergeben kann. Zudem können sie eine wirksame Behandlung einsetzen, um Änderungen hervorzurufen. Eine weitere Möglichkeit besteht darin, zusätzliche Messungen mit einem Referenzstandard durchzuführen und die Veränderungen, welche beide Tests anzeigen, miteinander zu vergleichen.

Eine wissenschaftliche Studie, welche die Sensitivität eines Tests untersucht, muss valide sein. In welchem Maß diese Anforderung erfüllt ist, untersucht der Therapeut anhand von Leitfragen, deren Inhalt davon abhängt, ob die Wissenschaftler die Sensitivität mit oder ohne Referenzstandard überprüft haben.

Ob die Evidenz praxisrelevant ist, zeigt sich in den statistischen Größen: bei der Validierung des Tests ohne Referenzstandard in Form der Effektgröße mit einem Mindestwert von 0,5 und bei der Verwendung eines Referenzstandards in Form des Korrelationskoeffizienten, dessen Absolutwert mindestens 0,4 betragen sollte. In beiden Fällen muss das Ergebnis statistisch signifikant sein ($p \leq 0,05$).

Zudem beurteilt der Therapeut die Anwendbarkeit des Tests. Dazu beantwortet er Leitfragen, beispielsweise, ob sein Patient den Test überhaupt durchführen würde und ob das Testergebnis für den Patienten nützlich wäre.

Kapitel 18 – Wirksamkeit einer Therapie

Zusammenfassung

Wirksamkeitsstudien überprüfen die Wirksamkeit einer Intervention im Vergleich zu einer anderen Behandlung, z. B. einer konventionellen Behandlung, einer Placebobehandlung oder keiner Behandlung.

Unter den Primärstudien bringt die randomisierte kontrollierte Doppel- bzw. Dreifachblindstudie die höchste Evidenz, sofern sie auch anderen wissenschaftlichen Gütekriterien genügt. Die Kriterien lassen sich mit Hilfe von Leitfragen überprüfen, z. B. ob die Gruppen ausreichend groß waren und ob eine Intention-to-Treat-Analyse durchgeführt wurde.

Ferner muss der Therapeut herausfinden, wie deutlich sich die Ergebnisse der Interventionsgruppe von denen der Vergleichs- oder Kontrollgruppe unterscheiden. Handelt es sich um dichotome Daten, so dienen Kenngrößen wie das RR, die ABI und NNT_B und deren Konfidenzintervalle zur Beurteilung. Bei Zielgrößen mit mehr als 2 Merkmalsausprägungen sind es v. a. die MCID und Effektgröße. Zudem ist darauf zu achten, ob die Unterschiede statistisch signifikant sind.

Abschließend beurteilt der Therapeut, ob sich die Evidenz bei seinem Patienten anwenden lässt. Dazu schätzt er ab, wie sehr sich der Patient von denjenigen in der Studie unterscheidet und passt ggf. die NNT_B für seinen Patienten an. Zudem berücksichtigt er die Wertvorstellungen und Präferenzen des Patienten und überlegt, ob er die Behandlung auf dieselbe Art wie in der Studie durchführen kann.

Kapitel 19 – Prävention und Hilfsmittel

Zusammenfassung

Ziele der Prävention sind, Krankheiten und gesundheitliche Schäden zu verhindern (Primärprävention), sie frühzeitig zu identifizieren und eine Frühtherapie einzuleiten (Sekundärprävention) und bei bereits erfolgter Krankheit deren Rückfall oder Verschlimmerung vorzubeugen (Tertiärprävention). Diese 3 Präventionsarten bzw. deren Ziele erfordern in der EBP folgende Strategien:

- Studien über die Früherkennung von Krankheiten werden in der EBP wie diagnostische Tests behandelt (► Kap. 12).
- Studien zu den anderen Präventionsarten/Zielen der Prävention werden wie Wirksamkeitsstudien ausgewertet (► Kap. 18).

Hilfsmittel dienen dazu, Alltagsaktivitäten und die Arbeit besser und gesundheitsschonender zu bewältigen. Besonders bei komplexeren und teureren Hilfsmitteln sollte der Therapeut nach wissenschaftlichen Studien suchen und deren Evidenz hinsichtlich der Validität, Relevanz und Anwendbarkeit wie bei den Wirksamkeitsstudien (► Kap. 18) beurteilen.

Bei einfacheren und gängigen Hilfsmitteln und/oder solchen, über die es keine wissenschaftliche Literatur gibt, sollte der Patient das Hilfsmittel direkt ausprobieren. Ist der Nutzen für den Patienten nicht sofort ersichtlich, eignen sich die systematischen Beobachtungen oder das Single-subject research design, um Evidenz über die Wirksamkeit zu erhalten.

Kapitel 20 – Nebenwirkungen

Zusammenfassung

Viele medizinische Behandlungen, aber auch manche ergo- oder physiotherapeutische Interventionen bringen unerwünschte Nebenwirkungen mit sich. Wissenschaftliche Studien überprüfen das Ausmaß, mit welchem potentielle Nebenwirkungen auftreten können. Dafür eignen sich bei den Primärstudien besonders RCTs. Sie vergleichen die Ergebnisse einer Gruppe, die der Exposition (Intervention) ausgesetzt sind, mit denjenigen einer Kontrollgruppe.

Weil sich die Studien hinsichtlich ihrer wissenschaftlichen Güte unterscheiden, muss der Therapeut deren Glaubwürdigkeit (Validität) einschätzen. Das erfolgt anhand bestimmter Leitfragen, die z.B.

nach der Gleichartigkeit der Studiengruppen, nach der Verblindung und nach dem logischen Zusammenhang zwischen Ursache und Wirkung fragen.

Als Nächstes muss der Therapeut herausfinden, wie deutlich die Nebenwirkung durch die Exposition zustande kommt. Dazu gibt es mehrere Kenngrößen wie das RR, die RO, RRI, ARI, und die NNT_H , welche im Kapitel Ätiologie genauer erklärt sind.

Zum Schluss beurteilt der Therapeut mit Hilfe von Leitfragen, inwiefern sich die Evidenz bei dem aktuellen Patienten in die Praxis umsetzen lässt. Dazu schätzt er ab, wie sehr sich der Patient von denjenigen in der Studie unterscheidet und passt ggf. die NNT_H für seinen Patienten in Form der NNT_{H-ind} an. Zudem fragt er danach, welche praktische Konsequenzen es für den Patienten hätte, würde er die Exposition vermeiden, und ob der Patient dazu bereit ist.

Kapitel 21 – Übersichtsartikel

Zusammenfassung

Übersichtsartikel vermitteln einen Überblick über die wissenschaftliche Literatur zu einem Thema. Es lassen sich 3 verschiedene Arten an Übersichtsartikeln unterscheiden: narrative Übersichtsartikel, systematische Übersichtsartikel und Meta-Analysen. Letztere 2 Arten sind für die EBP sehr wertvoll, narrative Übersichtsartikel stehen dagegen auf niedriger Evidenzstufe.

Hinsichtlich der wissenschaftlichen Güte unterscheiden sich jedoch auch z.B. systematische Übersichtsartikel untereinander. Sie lässt sich mithilfe von Leitfragen überprüfen, z.B. ob die Kriterien, nach denen Artikel ein- bzw. ausgeschlossen wurden, klar beschrieben und geeignet waren und ob die Validität der eingeschlossenen Studien bewertet wurde.

Außerdem muss der Therapeut die Relevanz der Ergebnisse für die Praxis bewerten. Dazu beantwortet er die Leitfragen, ob die Ergebnisse der eingeschlossenen Studien sich ähnelten, wie groß der Effekt der Intervention war, wie präzise die Abschätzung der Ergebnisse war. Die Kenngrößen sind dabei dieselben wie in den Primärstudien, z.B. die NNT in Wirksamkeitsstudien oder die Likelihood Ratio bei Tests. In Meta-Analysen, welche eigene Berechnungen durchführen, finden sich häufig z.B. gewichtete mittlere Effektgrößen.

Zum Schluss beurteilt der Therapeut mithilfe von Leitfragen, ob sich die Evidenz bei seinem Patienten anwenden lässt. Dazu schätzt er z.B. ab, ob sich die Ergebnisse der Studie auf seinen Patienten übertragen lassen und ob der Nutzen die Nebenwirkungen und Kosten wert ist.

Kapitel 22 – Leitlinien

Zusammenfassung

Leitlinien sind wie Wegweiser oder Entscheidungskorridore. Sie geben den medizinisch-therapeutischen Fachpersonen Hilfestellungen für eine sinnvolle Vorgehensweise bei bestimmten Gesundheitsbeeinträchtigungen ihrer Patienten. Leitlinien beinhalten nicht nur Behandlungsmethoden, sondern auch Themenbereiche wie die Prävention, (diagnostische) Tests, Prognose, Nebenwirkungen und Ökonomie.

Im Gegensatz zu Richtlinien sind Leitlinien rechtlich nicht bindend. Sie werden von Arbeitsgemeinschaften bzw. Expertengremien erarbeitet und lassen sich in speziellen Datenbanken im Internet finden. Leitlinien stehen auf unterschiedlichem Entwicklungsstand und wissenschaftlichem Niveau. Beispielsweise entwickelt die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) Leitlinien in einem 3-stufigen Prozess. Sie kennzeichnet die unterste Stufe mit geringstem Evidenzniveau mit S1, die mittlere mit S2 und die oberste mit S3. Es gibt einige Instrumente, um Leitlinien zu entwickeln und zu beurteilen, z.B. das Deutsche Leitlinien-Bewertungs-Instrument (DELB), welches Beurteilungskriterien zur methodischen Qualität und Praktikabilität von Leitlinien

enthält und frei zugänglich ist. Es ermöglicht Therapeuten und Therapeutinnen, selbst eine Beurteilung einer Leitlinie vorzunehmen.

Zusätzlich zu diesen Instrumenten bieten sich noch Leitfragen an, um die Übertragbarkeit der Evidenz auf den eigenen Patienten zu überprüfen, z.B. ob das primäre Ziel der Leitlinie mit dem eigenen Ziel übereinstimmt.

Kapitel 23 – Evidenzbasierte Praxis im Arbeitsalltag

Zusammenfassung

Verglichen mit dem herkömmlichen Therapiealltag ist die Integration der EBP mit einem erhöhten Zeitaufwand und zusätzlichen Kosten verbunden. Die EBP ist zudem eine intellektuelle Herausforderung und kann z.B. bei Vorgesetzten oder der Klinikleitung auf Unverständnis und Widerstand stoßen. Deshalb sind einige Vorüberlegungen und Strategien hilfreich, um die EBP in den Arbeitsalltag zu integrieren. Dazu gehört, Überzeugungsarbeit zu leisten, Kosten zu budgetieren (personeller Aufwand, Schulung, Beratung, Lizenzen), die notwendige Infrastruktur bereitzustellen sowie die personellen Voraussetzungen (z.B. Zuständigkeiten, Kenntniserwerb, Arbeitsgruppen) und zeitlichen Freiraum zu schaffen. Nicht zuletzt sind diplomatische Fähigkeiten erforderlich, um Hindernisse zu überwinden.

Zum Wohle der Patienten und der eigenen Professionalität lohnt es sich, die EBP zu etablieren. Trotzdem gibt es Grenzen bei der EBP. Sie resultieren erstens aus der niedrigen Qualität mancher Studien. Zweitens gibt es Studien, welche nicht das »gewünschte« Ergebnis erbracht haben und daher unveröffentlicht bleiben. Sie verzerren das wissenschaftliche Gesamtbild über die betreffende Intervention. Drittens beziehen sich wissenschaftliche Aussagen auf ein Patientenkollektiv. Diese Aussagen müssen nicht immer auf den Patienten in der Praxis zutreffen.

In letzteren Nachteil der EBP kann eine andere Ebene des evidenzbasierten Arbeitens korrigierend eingreifen: die systematischen Beobachtungen am Patienten. Diese Ebene überprüft die Wirksamkeit einer Intervention am aktuellen Patienten. Ergänzt werden diese beiden Ebenen durch die reflektierte Praxis, welche bei der Lösung weiterer Probleme in der Behandlung des Patienten hilft.

Single-subject research design

Single-subject research design – das steht für höchste Evidenz, wenn es um eine Studie mit einem einzelnen Patienten geht. Zugabe, diese wissenschaftlich anerkannte Methode ist für den Therapiealltag meistens zu aufwändig, doch beispielsweise beim Ausprobieren von Hilfsmitteln kann das Design sehr hilfreich sein.

Ziel des Single-subject research design ist, die Wirksamkeit einer Behandlung im Vergleich zu keiner Behandlung, einem Placebo oder zu einer anderen Behandlung an einem Patienten nachzuweisen. Im Vergleich zu den systematischen Beobachtungen am Patienten (► Buch:¹ Kap. 3) erbringt das Single-subject research design zwar höhere Evidenz, ist aber auch aufwändiger, denn die Studie geht normalerweise über einen längeren Zeitraum und umfasst mehr Messungen.

1 Einordnung in den wissenschaftlichen Kontext

Das Single-subject research design, auch N-gleich-1-Studie bzw. im Englischen N-of-1-study oder Single-case study genannt, bildet eine **wissenschaftlich anerkannte** Art von Studien an einzelnen Patienten und lässt sich als wissenschaftliche Forschungsarbeit publizieren.

► Single-subject research design bedeutet nicht dasselbe wie Fallstudie.

Eine **Fallstudie** beschreibt einen Patienten und dessen Reaktion auf eine Behandlung. Der Therapeut oder Arzt verzichtet jedoch darauf, die unabhängige Variable (► Buch: Kap. 3.2.1) zu manipulieren, z. B. in Form eines zwischenzeitlichen Therapieentzugs. Es fehlt eine Referenz, eine Kontrolle. Daher weiß man letztendlich nicht, ob die Therapie, ein Placeboeffekt oder eine Spontanheilung die Ursache für die Verbesserungen des Zustandes war.

Beim **Single-subject research design** dient in der Regel die Versuchsperson als **eigene Kontrolle**. Nur in Ausnahmefällen, die unten erläutert sind, ziehen die Studien wenige weitere Versuchspersonen hinzu. Aufgrund der Kontrolle stehen diese Studien auf einer höheren wissenschaftlichen Stufe als die Fallstudien. Dementsprechend ist die Güte der Evidenz höher als bei Fallstudien, jedoch niedriger als bei wissenschaftlichen Studien mit Patientengruppen.

Single-subject research design und Fallstudien untersuchen meist nur eine Person, deshalb lassen sich die Ergebnisse **nicht generalisieren**, also ohne weitere Studien mit Patientengruppen nicht auf andere Personen übertragen. Oder wissenschaftlich ausgedrückt: Die **externe Validität** ist nicht gewährleistet. Die Studien können aber als Basis dienen, um danach größere klinische Studien mit einer umfangreicheren Patientengruppe durchzuführen. Damit ließe sich schließlich auch die externe Validität überprüfen.

Studien des Typs Single-subject research design sind nicht alle gleich gestaltet, sondern zeichnen sich durch unterschiedliche Designs aus. Die wichtigsten sind:

- AB Design,
- Withdrawal Design,
- Multiple Baseline Design,
- Alternating Treatments Design.

Die folgenden Abschnitte zeigen zunächst die Gemeinsamkeiten der verschiedenen Typen auf, danach die Charakteristika der einzelnen Designs.

2 Grundschemata des Single-subject research design

Allen Single-subject research design gemeinsam sind die systematischen, wiederholten Messungen der abhängigen Variablen während einer oder mehrerer Anfangsphasen (Baselines) und Interventionsphasen.² Zur Bestimmung einer Baseline misst der Therapeut in mindestens drei therapeutischen Sitzungen die abhängige Variable, bevor die Intervention beginnt bzw. nach einer Unterbrechung wieder beginnt. Während der Interventionsphase(n) erfasst er ebenfalls mehrmals die abhängige Variable in verschiedenen Therapiestunden (Backman et al. 1997). Wenn möglich, wissen der Therapeut bzw. Arzt und der Patient nicht, in welcher Phase sich der Patient gerade befindet, d. h. ob eine Intervention angewendet wird und wenn ja, welche. In der Medizin – z. B. bei Medikamentengabe – ist diese **Verblindung** durch Gabe von Placebos häufig realisierbar, im Gegensatz zur Ergo- oder Physiotherapie. Allerdings kann vielleicht wenigstens der Therapeut, der die Messungen durchführt, verblindet sein. Das setzt voraus, dass der behandelnde Therapeut nicht dieselbe Person ist, die erfasst.

Grundsätzlich darf der Therapeut nur eine unabhängige Variable verändern, beispielsweise nur die Art der ergotherapeutischen Intervention (inklusive Placebo oder keine Intervention), aber nicht noch parallel dazu die medikamentöse Therapie. Außerdem muss die Baseline vor Einführung der Intervention stabil sein (Gonnella 1989). Um beurteilen zu können, ob die **Baseline stabil** ist, hieße das, dass es eigentlich mindestens 5 Messpunkte bräuchte, denn in den meisten klinischen Situationen schwanken die Werte von Tag zu Tag (Backman et al. 1997).

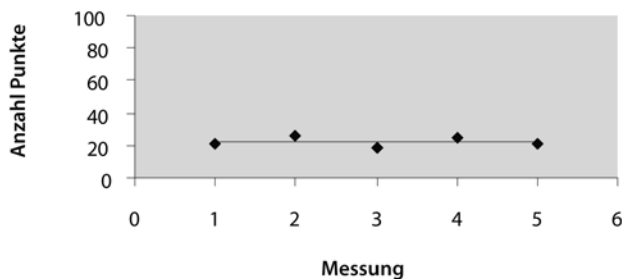
Was ist nun unter einer stabilen Baseline zu verstehen? Grundsätzlich lassen sich 2 Arten unterscheiden:

- Baseline mit konstantem Wert,
- Baseline mit konstanter Steigung.

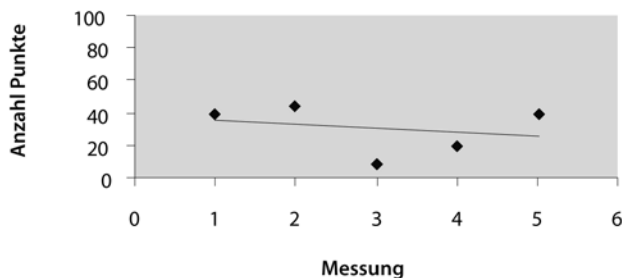
Es folgt die Erklärung der beiden Arten.

1 Verweise auf „Buch“ beziehen sich immer auf Mangold (2013): Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie, 2. Auflage

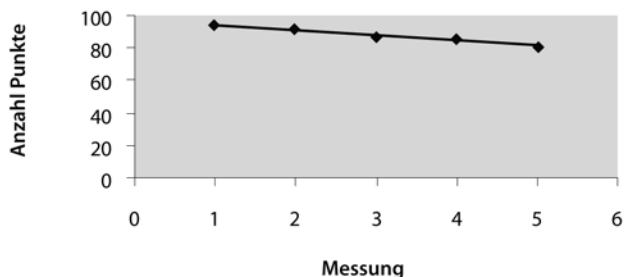
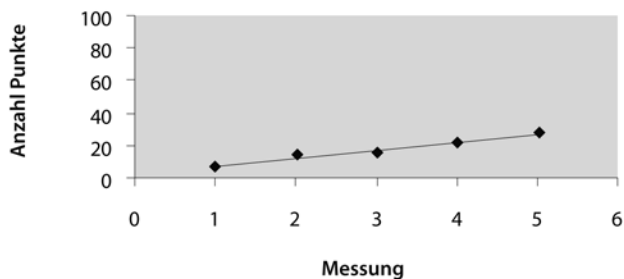
2 Das weiter unten beschriebene Alternating Treatments Design bildet allerdings eine Ausnahme, denn eine Baseline ist nicht zwangsläufig notwendig, jedoch empfehlenswert.



■ Abb. 1 Stabile Baseline mit konstantem Wert



■ Abb. 3 Instabile Baseline



■ Abb. 2a,b Stabile Baseline mit konstanter Steigung. a Kontinuierliche Steigerung der Punktzahl, b kontinuierliche Verminderung der Punktzahl

2.1 Arten stabiler Baselines

Baseline mit konstantem Wert

Bei dieser Art stabiler Baseline bleiben die Messpunkte auf dem gleichen Wert bzw. schwanken um einen mittleren Wert. ■ Abb. 1 zeigt als Beispiel 5 Baseline-Messungen, bei denen der Patient vor Einführung der Intervention durchschnittlich ungefähr 22,5 von 100 Punkten im Test, z. B. in einem Selbständigkeitsstatus, erreicht hat. Die Werte schwanken um diesen Wert. Die rechnerisch ermittelte waagerechte Linie durch die Punkte bedeutet, dass eine stabile Baseline mit konstantem Wert vorliegt.

Baseline mit konstanter Steigung

Bei der stabilen Baseline mit konstanter Steigung steigen bzw. fallen die Messpunkte kontinuierlich. Als Beispiel zeigen 5 Baseline-Messungen in ■ Abb. 2, dass der Patient vor Einführung der Intervention die Anzahl erreichter Punkte im Test kontinuierlich gesteigert (■ Abb. 2a) bzw. vermindert hat (■ Abb. 2b). Die Werte schwanken etwas um die eingezeichnete Ausgleichsgerade.

■ Abb. 3 zeigt als Vergleich ein Beispiel einer **instabilen** Baseline. Es lässt sich nicht erkennen, ob die Werte voraussichtlich weiter steigen würden, wie die letzten drei Messungen andeuten, oder ob es wieder zu einem großen Abfall kommen könnte wie zwischen der 2. und 3. Messung.

2.2 Vergleich der Baseline- und Interventionsphasen

Anhand des Verlaufs der Messresultate in den verschiedenen Phasen lassen sich eventuelle Unterschiede zwischen Baseline- und Interventionsphasen erkennen. Drei deutliche und wichtige Datenmuster sind folgende (Backman et al. 1997; eine ausführlichere Beschreibung möglicher Datenmuster findet sich in Wolery u. Harris 1982):

- Die Messwerte der Interventionsphase schwanken um einen anderen mittleren Wert als bei der Baseline, der **Level** hat sich also geändert (■ Abb. 4).

- Die **Steigung** ist eine andere. Sie kann steiler, weniger steil oder in umgekehrter Richtung erfolgen (■ Abb. 5).
- Die **Variabilität** (Streuung) der Daten ist verschieden. Das ist z. B. der Fall, wenn die Daten der Baseline sehr stark um einen Mittelwert schwanken, während sie bei der Intervention viel weniger von diesem Wert abweichen (■ Abb. 6).

Die **Auswertung** erfolgt meistens mithilfe einer visuellen Analyse der graphischen Darstellung der Daten. Sie betrachtet

- die Veränderungen der Levels der Daten in den einzelnen Phasen,
- den Anstieg beim Übergang von einer Phase zur nächsten,
- das Vorhandensein oder Fehlen eines Anstiegs bzw. Abfalls innerhalb einer Phase,
- die Änderungen der Variabilität in den verschiedenen Phasen sowie bei deren Übergang.

Auch mithilfe statistischer Methoden lassen sich die Daten auswerten, deren Erläuterungen hier zu weit führen würden. Eine genauere Beschreibung darüber findet sich in Bortz u. Döring (2006, S. 580–597).

3 Designs

Die ■ Abb. 4, ■ Abb. 5, ■ Abb. 6 zeigen jeweils eine Baseline und eine Interventionsphase. Es gibt jedoch verschiedene Varianten des Single-subject research design, also unterschiedliche Designs. Sie unterscheiden sich hinsichtlich der **Anzahl an Baseline- und Interventionsphasen**.

Bei der Charakterisierung ist es üblich, die Untersuchungsphasen mit den Buchstaben A und B zu kennzeichnen:

- **A steht für Baseline**, d. h. die Phase, in welcher der Patient keine Intervention, ein Placebo oder eine Standardtherapie erhält.
- **B bedeutet Interventionsphase** unter Anwendung derjenigen Intervention, deren Wirksamkeit Gegenstand der Untersuchung darstellt.

Die folgenden Abschnitte beschreiben 4 wichtige Arten des Single-subject research design.

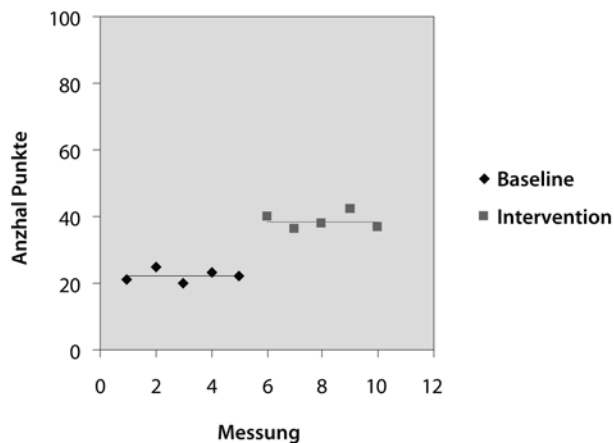


Abb. 4 Änderung des Levels. Die Abbildung zeigt 10 Messwerte (Messung 1-10) z. B. eines Handfunktionstests oder eines Selbsthilfestatus mit maximal 100 Punkten. Jeder Messwert stellt die Anzahl erreichter Punkte der betreffenden Messung dar. Die ersten 5 Werte bilden die Baseline, die weiteren Werte sind Messergebnisse in der Interventionsphase. Es ist deutlich zu erkennen, dass die Werte bei der Baseline um einen wesentlich niedrigeren Wert als bei der Anwendung der Intervention, z. B. Gebrauch eines Hilfsmittels, schwanken

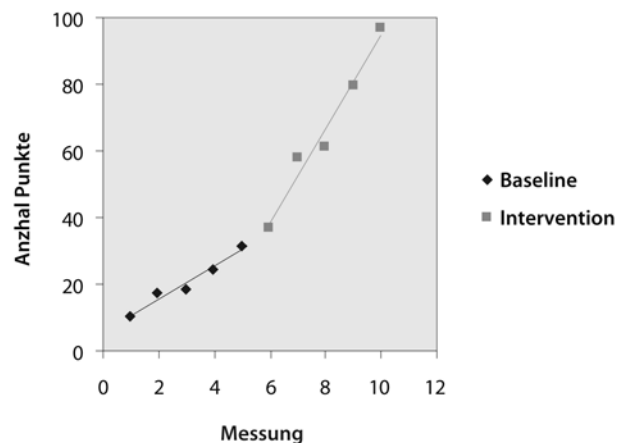


Abb. 5 Änderung der Steigung. Die ersten 5 Messwerte bilden die Baseline, bei der 6. und 10. Messung wurde die Intervention angewendet. Hier ist die Steigung bei Anwendung der Intervention höher als bei der Baseline. Sofern die zeitlichen Abstände zwischen den Messpunkten gleich sind, zeigt der Patient also schnellere Fortschritte als ohne Intervention

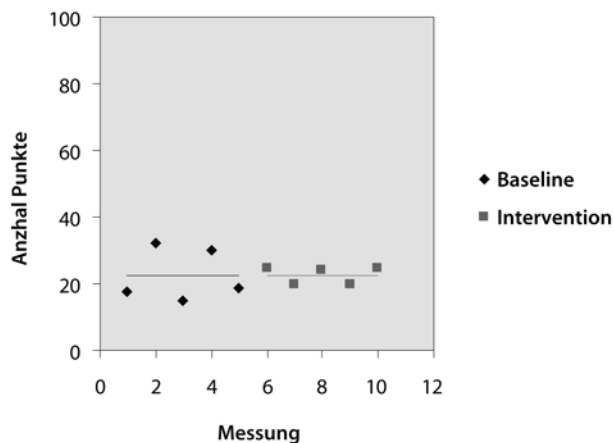


Abb. 6 Änderung der Variabilität. Die ersten 5 Messwerte bilden die Baseline, bei der 6.-10. Messung wurde die Intervention angewendet. Bei Anwendung der Intervention schwanken die Werte weniger als bei der Baseline, d. h. die Variabilität ist niedriger



Abb. 7 Single-subject research design

3.1 AB Design

Das AB Design ist das grundlegende und einfachste Single-subject research design. Es besteht aus einer Baseline (Phase A) und einer Interventionsphase (Phase B). Die Intervention darf, wie oben erwähnt, erst zur Anwendung kommen, wenn die Baseline stabil ist (Abb. 1, Abb. 2). Der Vergleich des Datenverlaufs der Baseline- und der Interventionsphase zeigt, ob z. B. Veränderungen hinsichtlich des Levels, der Steigung und/oder der Variabilität vorhanden sind (Abb. 4, Abb. 5, Abb. 6).

Das AB Design hat den Nachteil, dass – trotz stabiler Baseline – ein natürlicher Erholungsprozess nicht auszuschließen ist.

Praxistipp

Das Single-subject research design ist im Rahmen des normalen Therapiealltags zur Überprüfung der Wirkung von Therapiekonzepten wie dem Bobath-Konzept zu aufwändig. Aber es eignet sich dafür herauszufinden, welches Hilfsmittel für den Patienten geeignet ist, wie immer besonders dann, wenn Unsicherheiten bestehen. Im einfachsten Fall – dem AB Design – testet man die Funktion (z. B. Handfunktion) mehrere Male ohne das Hilfsmittel, dann mit. Das kann auch in einer einzigen Therapiesitzung geschehen.

3.2 Withdrawal Design

Das Withdrawal Design ist eine Erweiterung und Verbesserung des AB Design. Withdrawal (engl.) heißt Entzug. Wie beim AB Design folgt nach der anfänglichen Baseline (Phase A) die Interventionsphase mit entsprechenden Messwiederholungen (Pha-

se B). Danach schließt nochmals eine Baseline ohne Intervention an (Phase A). Die Intervention wird also in der Schlussphase wieder weggenommen, daher die Bezeichnung. Aufgrund der Phasen heißt das Withdrawal Design auch ABA Design.

Im Vergleich zum einfachen AB Design bietet das ABA Design folgenden Vorteil: Die abschließende Baseline identifiziert Funktionsverbesserungen, die nicht auf die Intervention, sondern z. B. auf natürliche Erholungsprozesse zurückzuführen sind. Wenn also die gemessene Funktion des Patienten vor dem Beginn und nach Absetzen der Intervention auf einem niedrigeren Level als während der Interventionsphase liegt, so erhärtet das die Annahme, dass der Effekt auf der Intervention beruht, besser als das AB Design.

Fallen die Werte bei der 2. Baseline nicht wieder auf die Anfangswerte (1. Baseline) zurück, so muss eine gefährdete interne Validität (mangelnde methodische Qualität) als mögliche Erklärung für die Änderungen während der Interventionsphase angenommen werden (Backman et al. 1997). Diese Forderung bringt allerdings mit sich, dass man die Wirksamkeit nachhaltig wirkender Interventionen, bei denen der Patient also auch nach deren Absetzen noch bessere Funktionen aufweist als vor deren Anwendung, mit diesem Design nicht nachweisen kann. Geeignet ist dieses Design aber beispielsweise für Hilfsmittel oder Schienen, deren möglicher Nutzen nur unmittelbar bei deren Anwendung vorliegt.

Das ABA Design lässt sich durch zusätzliche Phasen erweitern, z. B. ABABAB, um den Effekt der Intervention noch besser zu überprüfen und zu erhärten. Außerdem ist es auch erlaubt, andere Interventionen einzuführen und wieder zu entziehen, z. B. ABACA.

Abschließend ist ein wichtiger **ethischer Aspekt** des Withdrawal Design zu erwähnen: Der Entzug der Intervention muss ethisch vertretbar sein. Wenn sich beispielsweise ein Patient in der Akutphase eines Schlaganfalls befindet, sollte er fortlaufend Therapie erhalten. Ethisch vertretbar wäre aber, wenn **zusätzlich** zu der üblichen therapeutischen Versorgung eine neue Intervention oder ein Hilfsmittel mithilfe des Withdrawal Design erprobt würde.

3.3 Multiple Baseline Design

Das Multiple Baseline Design »spielt« mit der Dauer der Baselines. Man unterscheidet folgende Arten:

- Die Variation der Dauer der Baselinephase
- bei verschiedenen Probanden (Multiple baseline across subjects),
- in unterschiedlichen Settings (Multiple baseline across settings),
- bei verschiedenen Verhaltensweisen/Aufgabenbewältigungen (Multiple baseline across behavior).

Variation der Dauer der Baselinephase bei verschiedenen Probanden

Diese Art des Multiple Baseline Design erfordert **mehrere Patienten**. Der Therapeut variiert bei dieser Methode die **Länge der Baseline bei verschiedenen Patienten**. Jeder Patient verharret also unterschiedlich lang in der Baselinephase. Dadurch sollen sich Spontaneffekte ohne Therapie zeigen, die vielleicht bei einer langen Baselinephase, nicht aber bei einer kurzen auftreten. Die Wirksamkeit der Intervention wird dann ersichtlich, wenn die Werte während der Baselinephase – ob sie kurz ist oder lang – niedrig bleiben und jeweils dann ansteigen, sobald die Intervention anfängt.

Variation der Dauer der Baselinephase in unterschiedlichen Settings

Diese Form lässt sich auch bei einzelnen Patienten anwenden. Dabei variiert man die **Dauer der Baseline in verschiedenen Settings**. Die Erklärung erfolgt an einem Beispiel.

Beispiel: Dauer der Baseline in verschiedenen Settings

Bestehen die unterschiedlichen Settings aus dem Therapieraum, der eigenen Wohnung und dem Arbeitsplatz, so misst man zunächst die Baselinewerte in allen Settings. Dann wendet man die Intervention, z. B. ein Kommunikationshilfsmittel, vorerst nur in einem Setting an, z. B. am Arbeitsplatz. Ist die Intervention wirksam, so zeigt sich zwar am Arbeitsplatz, nicht aber in den anderen Settings eine Verbesserung, da sie sich ja noch nicht in der Interventionsphase befinden. Sobald die Intervention beim nächsten Setting angewendet wird, sollten auch dort Verbesserungen zu bemerken sein.

Wichtig ist bei diesem Design, dass die unabhängige Variable sich nicht global auf die verschiedenen Settings auswirkt. Trainiert die Intervention beispielsweise die aktive Sprechfunktion des Patienten, so wirkt sich das gleichzeitig auf alle Settings aus. Daher wäre diese Art des Multiple Baseline Design in diesem Fall nicht geeignet. Stellt die Intervention dagegen ein Hilfsmittel dar, so zeigt sich dessen Effekt nur in demjenigen Setting, in welchem der Patient es anwendet.

Variation der Dauer der Baselinephase bei verschiedenen Verhaltensweisen/Aufgabenbewältigungen

Hier untersucht der Therapeut anhand mehrerer voneinander **unabhängiger Aufgabenstellungen**, wie sich eine Intervention auswirkt. Die Aufgaben oder Aufgabengruppen werden unterschiedlich lang in der Baseline gehalten. Das Prinzip ist dasselbe wie bei der vorhergehenden Form (Variation der Dauer der Baselinephase in unterschiedlichen Settings).

Mit diesem Design untersuchten beispielsweise Mohr et al. (1999) die Wirkung eines elektronischen Hilfsmittels als externes Gedächtnis und Erinnerungshilfe bei Patienten mit einem starken amnestischen Syndrom. Sie wählten dazu 6 verschiedene Aufgaben aus dem Alltag aus, welche sie in 3 Aufgabengruppen einteilten.

Wie bei der vorhergehenden Art des Multiple Baseline Design ist wichtig, dass die Intervention die Bewältigung der Aufgaben unabhängig voneinander beeinflusst. Wenn die Intervention Grundlagenfähigkeiten für die Bewältigung aller Aufgaben trainiert, so verbessert sich auch die Bewältigung derjenigen Aufgaben, welche sich noch in der Baseline befinden.

3.4 Alternating Treatments Design

Diese Art des Multiple Baseline Design dient dazu, die relative Wirksamkeit mindestens zweier Interventionen bei einem Patienten zu untersuchen. Die verschiedenen Interventionen wechseln schnell in zufälliger (randomisierter) Reihenfolge ab (z. B. BDCCDBCBD). Vorteilhaft ist zudem, die Serie durch eine oder mehrere Baselines ohne Intervention zu ergänzen (z. B. ABDC-CDBCBD).

Der Unterschied zum AB Design besteht in der zeitlichen Randomisierung der Interventionen. Dadurch möchte man Effekte, die aufgrund der Reihenfolge auftreten könnten, neutralisieren.

Literatur

- Backman CL, Harris SR, Chisholm JM, Monette AD (1997) Single-subject research in rehabilitation: A review of studies using AB, withdrawal, multiple baseline, and alternating treatments design. *Arch Phys Med Rehabil* 78(10):1145-1153
- Bortz J, Döring N (2006) Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, 4. Aufl. Springer, Heidelberg
- Gonnella C (1989) Single subject experimental paradigm as a clinical decision tool. *Phys Ther* 69(7):601-609
- Mohr G, Kraeber A, Jochum I (1999) The effects of an electronic memory aid on the orientation of amnesic patients: A single case study. *Neurologie und Rehabilitation* 5(5):280-284
- Wolery M, Harris SR (1982) Interpreting results of single-subject research designs. *Phys Ther* 62(4):445-452

Arten wissenschaftlicher Publikationen

Randomisiert kontrollierte Studie – häufig zitiert und sehr geschätzt. Zu Recht, denn sie befindet sich aufgrund ihrer hohen Güte ganz oben in der Rangfolge wissenschaftlicher Studien, wenn auch nicht ganz alleine. Randomisiert kontrollierte Studien stehen aber nicht immer zur Verfügung, denn sie eignen sich nicht für jede Fragestellung und lassen sich in der Forschung nicht immer realisieren. Aber es existieren ja noch andere Studiendesigns. Welche sind das, was verbirgt sich dahinter und welches wissenschaftliche Niveau erreichen sie?

Studie ist nicht gleich Studie. Beispielsweise gibt es Untersuchungen an einzelnen Patienten oder Gruppen, Studien mit und ohne Kontrollgruppe und solche, welche die Ergebnisse anderer Studien zusammenfassen. Diese eher groben Kategorien fächern sich noch feiner in verschiedene Studiendesigns mit unterschiedlicher Beweiskraft auf.

Bereits bei der Suche wissenschaftlicher Literatur ist darauf zu achten, möglichst gute Studien für die Fragestellung zu selektieren, denn die EBP fordert die beste zur Verfügung stehende Evidenz. Das setzt voraus, dass man zumindest die wichtigsten bzw. häufigsten Designs und deren wissenschaftliche Einstufung kennt.

1 Arten wissenschaftlicher Publikationen in der quantitativen Forschung

Wissenschaftlerinnen und Wissenschaftler verfügen über zwei Möglichkeiten, Daten zu sammeln und zu veröffentlichen:

- Sie sammeln Daten an Patienten und analysieren und interpretieren sie.
- Sie suchen bereits veröffentlichte und unveröffentlichte Studien und analysieren und interpretieren deren Ergebnisse.

Aufgrund dieser unterschiedlichen Vorgehensweisen lassen sich wissenschaftliche Studien in 2 grobe Kategorien einteilen:

1. **Wissenschaftliche Primärstudien** sind Forschungsarbeiten, welche systematisch gesammelte Daten von Patienten unter einer wissenschaftlichen Fragestellung präsentieren. Beispiele dafür sind randomisierte kontrollierte Studien, Kohortenstudien oder Fall-Kontroll-Studien.
2. **Integrative Publikationen** fassen die Ergebnisse der Primärstudien und evtl. anderer integrativer Publikationen zusammen. Hierzu zählen die systematischen Übersichtsartikel (systematic reviews), Meta-Analysen, Leitlinien und Health Technology Assessments (HTA).

Den Lesern und Leserinnen solcher Studien bietet sich der Vorteil, dass sie bei der 1. Art sehr detaillierte Informationen über eine einzelne Studie erhalten und bei der 2. einen Überblick über die wesentlichen Informationen aus verschiedenen Studien zu einer wissenschaftlichen Fragestellung bekommen.

Die Primärstudien und die integrativen Publikationen fächern sich, wie die aufgeführten Beispiele bereits andeuteten, noch feiner in verschiedene Studiendesigns auf. Diesen unterschiedlichen Studiendesigns wird unterschiedliche Beweiskraft zugesprochen. In der EBP gibt es für jeden Themenbereich eine

eigene Hierarchie der Studiendesigns hinsichtlich ihrer wissenschaftlichen Aussagekraft (als Beispiel ► Buch: Tab. 18.1).

Praxistipp

In vielen Zusammenfassungen der wissenschaftlichen Artikel oder bereits in deren Titeln steht, um welches Studiendesign es sich handelt.

1.1 Wissenschaftliche Primärstudien

Eine der wichtigsten bzw. wissenschaftlich sehr anerkannten Primärstudien ist die randomisierte kontrollierte Studie. In der Wissenschaft existieren aber noch viele andere Studiendesigns. Alle aufzuführen und zu beschreiben, würde hier den Rahmen sprengen. Es folgt deshalb eine Auswahl an häufigen und wichtigen Studienarten.

Randomisierte kontrollierte Studie (engl.: Randomized Controlled Trial)

Die randomisierte kontrollierte Studie untersucht die Wirksamkeit einer zu prüfenden therapeutischen Intervention. Sie wird auch randomisierte klinische Studie genannt, und im Englischen heißt sie Randomized Trial, Randomized Controlled Trial, Randomized Clinical Trial oder einfach RCT.

- **Die RCT stellt den Goldenen Standard dar, wenn es um die Überprüfung der Wirksamkeit therapeutischer Interventionen geht. In diesem Zusammenhang bedeutet Goldener Standard die höchste Qualität wissenschaftlicher Primärstudien mit besonders hoher Evidenz.**

Bei diesem Design teilen die Wissenschaftler die Patienten in mindestens zwei Gruppen auf, und zwar nach dem Zufallsprinzip, also **randomisiert**. Diese randomisierte Zuordnung erfolgt beispielsweise durch ein Losverfahren oder durch einen elektronischen Zufallsgenerator. Die Patienten der einen Gruppe erhalten diejenige Behandlung, welche im Zentrum des Interesses steht. Diese Gruppe heißt Behandlungs- oder Interventionsgruppe. Die Patienten der anderen Gruppe(n) erhalten keine Behandlung, ein Placebo oder eine bisher übliche (konventionelle) Methode. Solche Gruppen heißen Kontroll- oder Vergleichsgruppen.

Aufgrund des Vergleichs mit einer Kontrollgruppe ist diese Studienart **kontrolliert**, daher der Name dieses Designs. Die Kontrolle ist bei diesem Studiendesign besonders effektiv, weil personenbezogene Unterschiede zwischen Behandlungs- und Kontrollgruppe durch die Randomisierung klein gehalten werden. Sozioökonomische, biologische oder medizinische Eigenschaften (Alter, Schweregrad der Krankheit, Einflüsse durch das persönliche Umfeld etc.) verteilen sich aufgrund der Randomisierung relativ gleichmäßig über die Gruppen und neutralisieren

1 Verweise auf »Buch« beziehen sich immer auf Mangold (2013): Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie

ren sich dadurch. Der einzige Unterschied bzw. beeinflussende Faktor sollte somit die Intervention sein.

Beispiel RCT-Design

Gitlin et al. (2001) untersuchten die Wirksamkeit ergotherapeutischer Behandlungen zu Hause bei Patienten mit Demenz bezüglich der Verhaltensprobleme der Patienten, Selbstpflege und Funktionen im instrumentellen ADL-Bereich der Patienten. Zudem wollten sie herausfinden, ob sich die psychische Belastung des Pflegepersonals ändert. Dazu schätzten sie deren Verstimmungen ein. Sie ordneten die Familien von Patienten mit Demenz randomisiert zu der Interventionsgruppe (ergotherapeutische Behandlung und konventionelle Pflege zu Hause) und Kontrollgruppe (nur konventionelle Pflege zu Hause) zu. Die ergotherapeutische Intervention betrug 5x90 min und beinhaltete sowohl Training der Patienten als auch Anpassungen der physischen und sozialen Umgebung.

Zum RCT-Design ist Folgendes anzumerken: Die RCT ist die aussagekräftigste, aber nicht immer die machbarste und praktischste Art zu forschen (Turner-Stokes 1999). **Ethisch** ist es im Gesundheitsbereich kritisch, einem Patienten, der das Los für die Kontrollgruppe gezogen hat, eine Behandlung der Forschung zuliebe vorzuenthalten. Möglich und ethisch vertretbar ist diese Methode jedoch in 2 Fällen:

- Wenn die Therapie wie im Beispiel von Gitlin et al. (2001) nicht fest zum klinisch-therapeutischen Programm gehört und die Patienten grundsätzlich versorgt sind.
- Wenn beide Gruppen mit einer konventionellen Therapie behandelt werden und die Interventionsgruppe **zusätzlich** die zu überprüfende Intervention erhält. Die gesamte Therapiedauer sollte jedoch bei beiden Gruppen übereinstimmen, denn sonst könnten die Effekte auch einfach durch die zusätzliche Therapiezeit zustande gekommen sein.

Nichtrandomisierte kontrollierte Studie

Die nichtrandomisierte kontrollierte Studie unterscheidet sich von der zuvor beschriebenen RCT durch die fehlende Randomisierung. Stattdessen teilen die Forscher die Patienten z. B. abwechselungsweise der Interventions- und Kontrollgruppe zu. Dadurch sinkt die wissenschaftliche Güte der Studie.

Nichtrandomisierte kontrollierte Studien, auch einfach kontrollierte Studien genannt, stehen bezüglich der wissenschaftlichen Qualität auf derselben Stufe wie die nachfolgend beschriebenen Kohortenstudien mit Kontrollgruppe (Donner-Banzhoff 2000). Allerdings gilt zu berücksichtigen, dass es unter den nichtrandomisierten kontrollierten Studien unterschiedliche Qualitäten gibt:

- Am niedrigsten sind diejenigen Studien einzustufen, welche die Zuteilung nach einem personenbezogenen Faktor vorgenommen haben (ausgenommen Matching). Werden z. B. Personen abhängig von ihrem Stadtkreis, in welchem sie wohnen, der Interventions- oder Kontrollgruppe zugeordnet, so kann das die Studie verzerren, weil sich die Bewohner verschiedener Stadtkreise hinsichtlich ihres Bildungsstandes, sozialer Faktoren, Einkommen etc. unterscheiden können.
- Eine mittlere Position unter den nichtrandomisierten Studien nehmen solche ein, welche die Patienten beispielsweise abwechselungsweise der Interventions- und Kontrollgruppe zuordnen.
- Am höchsten einzustufen sind Studien, bei denen die Wissenschaftler gezielt dafür sorgen, dass Störvariablen (weitgehend) gleichmäßig auf die Gruppen verteilt werden. Dazu bilden sie Paare von Patienten, welche sich hinsichtlich wichtiger Eigenschaften (z. B. Alter, Schweregrad der Erkrankung) ähnlich sind, und teilen pro Paar einen Patienten der Interventionsgruppe und den anderen der Kontroll- bzw. Vergleichsgruppe zu. Dieses Verfahren heißt **Matching** oder Matched Samples.

Crossover-Studie (engl.: Crossover Trial; Crossover Study)

Die Patienten einer Crossover-Studie erhalten mindestens 2 Interventionen nacheinander oder 1 Intervention und keine Intervention (Kontrollphase). Die Reihenfolge wird bei allen Versuchsteilnehmern randomisiert. Untersuchen die Wissenschaftler beispielsweise, ob 1 Intervention besser als keine Intervention ist, so erhalten manche Patienten zuerst die Intervention und dann keine mehr. Bei den anderen Patienten ist es umgekehrt. Die Variabilität aufgrund unterschiedlicher individueller Eigenschaften reduziert sich durch dieses Versuchsdesign, da alle Probanden Teil jeder Interventions- bzw. Kontrollgruppe sind.

Die Crossover-Studie stellt eine Erweiterung der kontrollierten Studie dar. Ihr Evidenzniveau liegt daher zwischen dem einer RCT und einer kontrollierten Studie.

Kohortenstudie (engl.: Cohort Study)

Allgemein ausgedrückt ist eine **Kohorte** eine Gruppe von Personen mit einer oder mehreren gemeinsamen Eigenschaften. Im Zusammenhang der Kohortenstudien beziehen sich diese gemeinsamen Eigenschaften je nach Fragestellung beispielsweise auf ein Behandlungsprogramm, eine Diagnose bzw. Krankheit, das Vorhandensein eines potentiellen Risikofaktors oder auch z. B. auf die Zugehörigkeit zu einer sozialen Gruppe. Eine Kohortenstudie konnte z. B. untersuchen, ob Patienten mit einer Hemiplegie (1. Kohorte) häufiger ein Karpaltunnelsyndrom als Menschen ohne Hemiplegie (2. Kohorte) entwickeln.

Ein wichtiger Begriff in der EBP im Zusammenhang der Kohortenstudien ist die **Inception Cohort Study**, die **Anfangs- oder Beginn-Kohortenstudie** (inception = Anfang, Beginn), denn er taucht in den Rastern zur Beurteilung der Evidenzniveaus der verschiedenen Studienarten auf (als Beispiel: ► Buch: Tab. 18.1). Anfangs-Kohortenstudie bedeutet, dass die Patienten zu einem gleichmäßig frühen Zeitpunkt ihrer Krankheitsentwicklung in die Studie aufgenommen werden. Die Patienten haben dann alle ungefähr denselben Startpunkt. Diese Homogenität ist wichtig, denn beispielsweise Patienten im chronischen Stadium haben häufig andere Prognosen als im akuten Stadium. Die Studie muss nennen, um welchen Zeitpunkt es sich handelt, z. B. 2 Wochen nach Diagnosestellung.

Es gibt noch weitere Differenzierungen der Kohortenstudien, wie die folgenden Abschnitte beschreiben. Spezifische Leitfragen zur Beurteilung der Güte der Kohortenstudien finden sich im Anhang C der MEDDEV 2.7.1. (2009).

Kohortenstudien mit Kontrollgruppe (engl.: Cohort Design)

Kohortenstudien mit Kontrollgruppe vergleichen mindestens 2 Kohorten miteinander. Der Unterschied zur randomisierten kontrollierten Studie besteht darin, dass der Versuchsleiter die Probanden nicht per Zufallsverfahren den 2 Gruppen zuteilt, denn die Eigenschaften, auf welchen die Zuordnung basiert, sind entweder bereits vorhanden (z. B. genetische Voraussetzungen, körperliche Eigenschaften, soziale oder berufliche Situation etc.) oder sie werden durch die Patienten beeinflusst, beispielsweise durch ihre individuelle Wahl, ob sie die angebotene Therapie wahrnehmen möchten oder nicht.

Häufig geht es in Kohortenstudien darum, das Auftreten einer Gesundheitsbeeinträchtigung oder die Weiterentwicklung einer Krankheit zu beobachten. Diese Studien heißen **Analytische Kohortenstudie** (engl.: Cohort Analytic Study). Es lassen sich 2 Formen unterscheiden:

- **Ersterkrankung:** Die 1. Art der analytischen Kohortenstudie untersucht diejenigen Faktoren, welche eine Gesundheitsbeeinträchtigung, z. B. Rückenschmerzen, verursachen konnten. In die Studie werden nur Menschen

aufgenommen, bei denen die Gesundheitsbeeinträchtigung, welche von Interesse ist, zum Zeitpunkt der Aufnahme in die Studie nicht nachweisbar ist. Die Studie ist also **prospektiv**. Die Versuchsteilnehmer werden 2 Kohorten zugeteilt: In der einen Kohorte befinden sich diejenigen, welche einer möglichen Ursache der Krankheit ausgesetzt sind (z. B. Arbeitsplatz mit hoher körperlicher Belastung), in der anderen diejenigen, welche der möglichen Ursache nicht ausgesetzt sind (Arbeitsplatz mit niedriger körperlicher Belastung). Nach einer bestimmten Zeit erfassen die Wissenschaftler die betreffende Zielgröße (z. B. Rückenschmerzen) und vergleichen die Ergebnisse der beiden Kohorten miteinander.

- **Weiterentwicklung einer Krankheit:** Die 2. Art der analytischen Kohortenstudie untersucht Patienten mit einer bereits bestehenden Krankheit oder Behinderung. Der positive oder negative Zustand, welcher im Mittelpunkt des Interesses steht und sich entwickeln könnte, ist zum Zeitpunkt der Aufnahme in die Studie noch nicht feststellbar. Die Studie ist also **prospektiv**. Das Ziel der Studie ist herauszufinden, inwiefern ein prognostischer Faktor (Merkmal, durch welches der Krankheitsverlauf besser abgeschätzt werden kann, z. B. Alter, Allgemeinzustand, Schweregrad der Erkrankung) mit einer bestimmten Krankheitsentwicklung zusammenhängt. Dazu teilt sie die Patienten hinsichtlich des prognostischen Faktors in verschiedene Kohorten ein. Wenn Wissenschaftler als prognostischen Faktor das Alter untersuchen, so teilen sie die Patienten beispielsweise in 2 Altersgruppen, also 2 Kohorten, auf. Die Wissenschaftler beobachten die Kohorten jeweils für eine bestimmte Zeit und vergleichen die Ergebnisse der beiden Kohorten miteinander. Wenn eine Studie als prognostischen Faktor eine Intervention wählt, so erhält die eine Kohorte die Intervention, deren Wirksamkeit untersucht werden soll, die andere Kohorte nicht. Zudem ist möglich, mit diesem Design die Wirksamkeit verschiedener Interventionen zu testen, indem jede Kohorte eine der Interventionen erhält.

Kohortenstudien können prospektiv oder retrospektiv sein (Donner-Banzhoff 2000). Die vorhergehend beschriebenen analytischen Kohortenstudien gehören zu den prospektiven Studien, weil bei Studienbeginn die Zielgröße, z. B. Krankheitsanzeichen, noch nicht feststellbar ist.

Retrospektive Kohortenstudien werten **im Nachhinein** Patientendaten aus, beispielsweise solche aus einer Patientenkartei oder Befragung der Patienten oder Angehörigen. Die Patienten gehören aufgrund ihrer Risikofaktoren oder prognostischen Faktoren zu einen oder anderen Kohorte. Die Wissenschaftler vergleichen die Ergebnisse der Kohorten miteinander. In der retrospektiven Kohortenstudie sind im Gegensatz zur prospektiven Studie die gesundheitlichen Beeinträchtigungen zum Zeitpunkt der Studiendurchführung bereits bekannt.

Wenn die Risikofaktoren bzw. prognostischen Faktoren von den Patienten selbst bzw. von deren Angehörigen im Nachhinein erfragt werden, kann es zu Fehleinschätzungen kommen. Dies führt zu Verzerrungen der Studienergebnisse, wodurch sich die Güte der Evidenz vermindert.

- **Kohortenstudien mit Kontrollgruppe gelten als wissenschaftlich wertvoll, da sie zu den kontrollierten klinischen Studien gehören. Allerdings stehen sie in der Rangfolge meistens hinter den randomisierten kontrollierten Studien.**

Der Nachteil der Kohortenstudie im Vergleich zur randomisierten kontrollierten Studie ist die geringere Wahrscheinlichkeit,

dass die Gruppen (Kohorten) gleiche Voraussetzungen haben. Dadurch können **Störvariablen**, welche nicht gleichmäßig über die Kohorten verteilt sind, die Studie erheblich verzerren.

Beispiel Verzerrung der Studienergebnisse durch Störvariablen

Wenn z. B. in einer pädiatrischen Studie die Kinder nicht zufallsbedingt in die Behandlungs- und Kontrollgruppe gelangen, sondern die Eltern der Kinder entscheiden, ob ihr Kind die Therapie bekommt oder nicht, so kann es sein, dass nicht nur die Therapie an sich, sondern auch Störvariablen das Resultat entscheidend mitbestimmen: Die Eltern der unbehandelten Kinder könnten z. B. das Krankheitsbild ihres Kindes unterschätzen und es daher auch privat weniger fördern als die Eltern der Kinder, die sich für die Therapie entscheiden.

Kohortenstudie ohne Kontrollgruppe

Dieses Studiendesign verzichtet auf eine Kontrollgruppe. Die Wissenschaftler wenden eine Intervention bei einer Kohorte an und beschreiben die Änderungen der Zielgröße. Solche Studien sind im Prinzip Anwendungsbeobachtungen. Da eine Kontrollgruppe fehlt, lassen sich Einflüsse wie Spontanheilung, Placeboeffekte etc. nicht abschätzen. Dementsprechend liegt ein geringeres wissenschaftliches Niveau als bei den kontrollierten klinischen Studien vor.

Alle-oder-Keiner-Studie (engl.: All or none)

Wenn eine Studie ohne Kontrollgruppe vorliegt, die einen sehr deutlichen Effekt der Intervention zeigt, so spricht man von einer Alle-oder-Keiner-Studie. Die Evidenz dieser Studie gilt in der EBM als hochwertig, wenn alle Patienten gestorben sind, bevor die Intervention verfügbar war, aber jetzt einige überleben oder wenn manche Patienten gestorben sind, bevor die Intervention verfügbar war, aber jetzt niemand mehr stirbt (Oxford Centre for Evidence-based Medicine 2009). Deutliche Effekte kann es beispielsweise bei der Anwendung von Antibiotika geben.

Wenn es um die Folgen einer schädlichen Exposition geht, so ist der Effekt sehr deutlich, »...wenn (fast) alle Personen unter Exposition ein bestimmtes Outcome erleiden – oder ohne Exposition (fast) keiner (Beispiel: Phkomelien nach Conterganexposition der Mütter in der Schwangerschaft)« (Ekkernkamp et al. 2003).

Fall-Kontroll-Studie (engl.: Case-Control Study):

Die Fall-Kontroll-Studie dient normalerweise dazu, mögliche Ursachen gesundheitlicher Beeinträchtigungen herauszufinden. Die Studie vergleicht Patienten, welche eine definierte Gesundheitsbeeinträchtigung haben (**Fälle**, engl.: cases), mit solchen ohne diese Beeinträchtigung (**Kontrollen**, engl.: controls). Fälle und Kontrollen sollen einander dabei hinsichtlich bestimmter Eigenschaften möglichst ähneln, z. B. bezüglich Alter, Geschlecht, Einnahme von Medikamenten, Begleiterkrankungen usw. Der Vergleich bezieht sich auf mögliche Risikofaktoren, welchen die Probanden beider Gruppen möglicherweise ausgesetzt waren. Besteht eine Verbindung zwischen einer solchen Exposition und einer gesundheitlichen Beeinträchtigung, so zeigt sich das durch einen höheren Anteil der Exponierten im Kollektiv der Fälle im Vergleich zum Kollektiv der Kontrollen.

Beispiel Fall-Kontroll-Studie

Untersuchen Wissenschaftler, ob eine Verbindung zwischen dem Karpaltunnelsyndrom (Erkrankung) und Arbeit am Computer (möglicher Risikofaktor) besteht, so sind die Fälle die Patienten mit dem Karpaltunnelsyndrom und die Kontrollen Patienten ohne Syndrom. Die Wissenschaftler ermitteln nun die Anzahl der Personen mit Computerarbeit (Exposition) und solche ohne Computerarbeit

(keine Exposition) sowohl im Kollektiv der Erkrankten als auch der Gesunden. Ein positiver Zusammenhang zwischen Computerarbeit und Karpaltunnelsyndrom besteht, wenn der Anteil Computerarbeiter in der Gruppe der Fälle bedeutend höher als in der Gruppe der Kontrollen, z. B. 56 % vs. 44 % ist.

Fall-Kontroll-Studien nehmen in der wissenschaftlichen Rangfolge eine mittlere Position ein. Spezifische Leitfragen zur Beurteilung der Güte der Fall-Kontroll-Studien finden sich im Anhang C der MEDDEV 2.7.1. (2009).

Querschnittstudie (engl.: Cross-Sectional Study)

Querschnittstudien nehmen einen »Schnappschuss« einer Population zu einem einzigen Zeitpunkt auf. Beispielsweise könnten Wissenschaftler die Lebensqualität von Menschen mit einer Sehbehinderung untersuchen. Dazu untersuchen sie eine Gruppe von Personen, die repräsentativ für die betreffende Population sind. Diese Personen werden in der Regel nicht noch einmal getestet bzw. befragt.

Das Evidenzniveau ist niedrig, es liegt auf derselben Stufe wie das der nachfolgend beschriebenen Fallserie.

Fallstudie (engl.: Case Report) und Fallserie (engl.: Case Series)

Die Fallstudie beschreibt einen Patienten hinsichtlich seiner Problemstellung und des Therapieverlaufs bzw. der Zielgrößen qualitativ und/oder quantitativ. Die Fallserie beschreibt dasselbe für eine Reihe von Patienten mit vergleichbarer gesundheitlicher Beeinträchtigung.

Gute Fallstudien beruhen auf systematisch erfassten Daten, z. B. auf systematischen Beobachtungen oder quantitativen Messungen. Ein Vergleich mit einer Kontrollgruppe fehlt jedoch in der Regel und es gibt **keine systematische Kontrolle** am Patienten selbst wie bei dem Single-subject research design (Internet-Link für Download: ► <http://extras.springer.com>).

Diese Arten von Studien sind aufgrund der fehlenden Kontrolle wissenschaftlich weniger wertvoll, wobei die Fallserien aufgrund der größeren Patientenzahl höher als die Fallstudien einzustufen sind. Sie dienen v. a. der Hypothesenbildung, d. h. die getroffenen Aussagen in der Fallstudie bzw. Fallserie können als Hypothese formuliert und anhand größerer Patientengruppen nochmals systematisch und kontrolliert überprüft werden. Spezifische Leitfragen zur Beurteilung der Güte der Fallserien finden sich im Anhang C der MEDDEV 2.7.1. (2009).

Single-subject research design (N-gleich-1-Studie, engl.: Single-subject study; N-of-1-study)

Zu diesen Studien an einzelnen Patienten, ► Internet-Link für Download: <http://extras.springer.com>. Sie lassen sich hinsichtlich der wissenschaftlichen Qualität und damit der Güte der Evidenz weit niedriger als die RCT einstufen, da sie die Wirksamkeit nur anhand eines Patienten überprüfen. Sie sind aber wertvoller als Fallstudien (► Abschn. »Fallstudie (engl.: Case Report) und Fallserie (engl.: Case Series)«), da der Patient die eigene Kontrolle darstellt.

Ökonomische Studien

Ökonomische Studien im Gesundheitswesen dienen dazu, die Kosten und den gesundheitlichen oder finanziellen Nutzen verschiedener medizinischer oder therapeutischer Interventionen zu vergleichen um herauszufinden, ob sich eine Intervention **aus ökonomischer Sicht lohnt** bzw. welche Intervention aus alternativen Vorschlägen zu bevorzugen ist. Die Bewertung der Kosten und Nutzen von Interventionen dienen als Entscheidungshilfe, nicht nur für die Fachdisziplinen des Gesundheitswesens selbst, sondern auch z. B. für **Gesetzgeber und Kostenträger** wie Krankenkassen und Unfallversicherungen.

Die wichtigsten Arten ökonomischer Studien sind:

- **Krankheitskostenstudie** (engl.: Cost-of-Illness Study, COI): Diese Art gehört **im weiteren Sinn** zu den ökonomischen Studien, da sie zwar die Kosten ermittelt, welche mit der medizinischen oder therapeutischen Versorgung verbunden sind, aber keine Aussage über den Nutzen trifft. Daher bietet sie keine Entscheidungshilfe darüber, welche Interventionsart hinsichtlich der finanziellen Investition und des Nutzens zu bevorzugen ist.
- **Kosten-Minimierungs-Analyse** (engl.: Cost-Minimization Analysis, CM): Diese Analyse vergleicht die Kosten zweier oder mehrerer Interventionen miteinander, welche **denselben Effekt** erbringen. Solche Kosten-Minimierungs-Analysen weisen häufig jedoch nicht selbst den Nutzen nach, sondern schließen an klinische Studien an, welche belegen, dass die primären Effekte der untersuchten Interventionen dieselben sind. Durch den Vergleich der Kosten lässt sich die günstigste Intervention identifizieren.
- Die **Kosten-Nutzen-Analyse** (engl.: Cost-Benefit Analysis, CBA), **Kosten-Effektivitäts-Analyse** (engl.: Cost-Effectiveness Analysis, CEA) und **Kosten-Nutzwert-Analyse** (engl.: Cost-Utility Analysis, CUA) ermitteln die Kosten sowie den monetären (d. h. Geld betreffenden) oder gesundheitlichen Nutzen verschiedener Interventionen. Indem die Studien Kosten und Nutzen miteinander verrechnen, lassen sich verschiedene Interventionen ökonomisch vergleichen. Zu ausführlichen Beschreibungen dieser Studienarten, ► Internet-Link für Download: <http://extras.springer.com>.

➤ **Ökonomische Studien im engeren Sinne berechnen nicht nur die Kosten, sondern berücksichtigen auch den Nutzen.**

Klinische Entscheidungsfindung (engl.: Clinical Decision Rule)²

Klinische Entscheidungsfindungen (engl.: Clinical Decision Rule, CDR) sind Studien, welche klinische Daten sammeln, um Algorithmen (Berechnungsmodelle) oder Punktesysteme aufzustellen, die zu einer prognostischen Schätzung oder einer diagnostischen Kategorie führen (Oxford Centre for Evidence-based Medicine 2009).

Beispiel klinische Entscheidungsfindung

Der Artikel von Atabaki et al. (2008) stellt eine klinische Entscheidungsfindung zur Verfügung, welche hilft zu entscheiden, ob sich ein Kind mit einem geringen Schädeltrauma einer durch die Strahlung belastenden und teuren kranialen Computertomographie für die frühe Entdeckung einer intrakraniellen Verletzung unterziehen soll. Um zu dieser klinischen Entscheidungsfindung zu kommen, sammelten die Wissenschaftler zunächst Daten junger Patienten mit einem geschlossenen Kopftrauma. Zu den Daten zählten beispielsweise Schwindel, sensorische Defizite, Schädelverletzung und Alter. Diese banden die Wissenschaftler in einen Entscheidungsbaum ein. Erfasst der praktizierende Kliniker nun dieselben Daten vom eigenen Patienten (z. B. Schwindel ja oder nein? Alter unter oder über 2 Jahren?), so sieht er anhand der Zweige des Entscheidungsbaumes der Studie, ob er eine kraniale Computertomographie anordnen soll oder ob er darauf verzichten kann.

2 Auch die Güte von Artikeln zur klinischen Entscheidungsfindung lassen sich mit Hilfe der EBM beurteilen. Interessierten Leserinnen und Lesern zu diesem Thema seien die zwei Artikel von Laupacis et al. (1997) und Stiell u. Wells (1999) empfohlen.

Ökologische Studie (engl.: Ecological Study)

Ökologische Studien im Gesundheitswesen sind Studien, welche sich auf Populationen auf der geographischen Ebene beziehen. Sie untersuchen beispielsweise den Zusammenhang zwischen physikalischen oder sozialen Faktoren und der Entwicklung bestimmter Krankheiten in **verschiedenen Ländern**.

Ökologische Studien sind in der EBM und EBP hinsichtlich ihrer Güte der Evidenz niedriger als z. B. Kohortenstudien und RCT eingestuft. Das liegt daran, dass die Daten sehr vieler Menschen mit sehr unterschiedlichen Voraussetzungen in die Statistik mit eingehen. Beispielsweise kommt es nicht nur auf das Land an, in dem die Menschen leben, sondern – je nach Fragestellung – auch z. B. auf den Bildungsstand oder die finanziellen Verhältnisse, die ja innerhalb der Population nicht einheitlich sind. Rückschlüsse auf einzelne Menschen zu ziehen, ist daher mit einer größeren Unsicherheit verbunden. Allerdings hängt das Ausmaß dieser Schwäche von der statistischen Methode ab, denn bestimmte Verfahrensweisen können die Daten differenzierter analysieren. Genauere Angaben hierzu finden sich in Blakely u. Woodward (2000). Zudem gibt es auch den Ansatz, die Studie mit Fall-Kontroll-Daten zu ergänzen, um die individuelle Ebene zu verstärken (Haneuse u. Wakefield 2008).

1.2 Integrative Publikationen

Systematischer Übersichtsartikel (engl.: Systematic Review)

Systematische Übersichtsartikel, auch systematische Reviews genannt, fassen so weit wie möglich alle wissenschaftlichen Informationen zu einer bestimmten Fragestellung zusammen. Sie interpretieren die Gesamtheit der verschiedenen Studienergebnisse kritisch, indem sie z. B. Widersprüche aufdecken und versuchen, diese zu klären. Sie zeigen Grenzen auf und bestimmen, ob eine untersuchte Intervention nur für eine Untergruppe gilt oder für eine größere Gruppe generalisierbar ist.

Systematisch bedeutet dabei, dass die Literatursuche nicht nur anhand der bekannten und viel verwendeten medizinischen Datenbank Medline, sondern umfassend mithilfe verschiedener Datenbanken und Suchmaschinen (z. B. PubMed, CINAHL, EMBASE) erfolgt. Zusätzlich verlesen die Autoren eines systematischen Übersichtsartikels verschiedene Zeitschriften von Hand und durchsuchen die Referenzlisten passender Artikel nach weiteren wichtigen Studien. Diese sorgfältige, systematische Literatursuche soll gewährleisten, dass sie keine wichtigen Studien übersehen.

Ferner ist unter systematisch zu verstehen, dass die Autoren und Autorinnen des Übersichtsartikels die gefundenen Studien hinsichtlich der wissenschaftlichen Qualität kritisch beurteilen und selektieren.

Systematische Übersichtsartikel, die z. B. über RCTs und Kohortenstudien mit Kontrollgruppen berichten, sind wissenschaftlich sehr wertvoll und besonders empfehlenswert für die EBP. Aber aufgepasst: Häufig ist einfach nur die Rede von »Übersichtsartikel«, und nicht alle stehen auf hohem Niveau. Vielmehr existieren verschiedene Arten von Übersichtsartikeln, die man für die Einstufung der Evidenz kennen sollte. Deshalb stellen die folgenden Abschnitte 2 weitere Formen von Übersichtsartikeln vor.

Narrativer Übersichtsartikel (engl.: Review)

Von den systematischen Übersichtsartikeln sind die narrativen Übersichtsartikel abzugrenzen, auch einfach Übersichtsartikel (Reviews) genannt. Übersichtsartikel fassen in kurzer und übersichtlicher Form das zusammen, was dem Verfasser zu einem bestimmten Thema oder zu einer speziellen Frage wichtig und

richtig scheint. Der Leser erhält so die Möglichkeit, sich einen Überblick über ein bestimmtes Thema zu verschaffen, der aber evtl. nicht vollständig ist. In der Regel werden Reviews von Experten und Expertinnen geschrieben, welche sich in dem beschriebenen Gebiet umfassend auskennen. Trotzdem besteht die Gefahr, dass die Autoren ihre **persönliche Meinung** zuungunsten einer neutralen Darstellung widerspiegeln und sie die Literatur nicht systematisch gesucht und bewertet haben. Es fehlt die Garantie, dass die wissenschaftlich hochwertigen Studien vollständig sind und ein höheres Gewicht erhalten als wissenschaftlich minderwertigere Studien. Übersichtsartikel stehen daher – im Gegensatz zu den systematischen Übersichtsartikeln – auf niedriger Evidenzstufe.

Wie erkennt man nun, um welche Art es sich handelt? Nun, das ist nicht schwer: Systematische Übersichtsartikel verfügen im Gegensatz zu den Übersichtsartikeln über die typische wissenschaftliche Struktur (Einleitung, Methoden etc.). Insbesondere führt der Methodenteil die systematischen Suchstrategien konkret auf.

Spezialform systematischer Übersichtsartikel: Meta-Analyse (engl.: Meta-analysis)

Die Meta-Analyse ist eine spezielle Form des systematischen Reviews. Auch sie vermittelt einen Überblick über die vorhandenen Studien unter einer bestimmten Fragestellung und klärt und interpretiert Widersprüche, die sich in der Literatur finden. Das Spezielle an der Meta-Analyse liegt darin, dass sie die experimentellen Daten der zugrunde liegenden klinischen Primärstudien statistisch auswertet. Weil dadurch insgesamt mehr Daten zusammenkommen als in einzelnen klinischen Studien, erhöht sich die Glaubwürdigkeit der Auswertungen und Aussagen.

Das Ziel einer Meta-Analyse ist also, mittels **Datensynthese** hohe Patientenzahlen zu erlangen, um genauere Aussagen z. B. über die Wirksamkeit einer Intervention treffen zu können. Um dieses Ziel zu erreichen, gehen die Wissenschaftler und Wissenschaftlerinnen folgendermaßen vor: Sie suchen klinische Studien, die zu einer bestimmten Fragestellung passen, von wissenschaftlich guter Qualität zeugen und deren Daten bezüglich der experimentellen Bedingungen vergleichbar sind. Diese Daten werten sie dann mithilfe spezieller Statistiken (Lipsey u. Wilson 2001, Deeks 2002) zusammenfassend aus.

➤ **Meta-Analysen sind aufgrund ihrer systematischen Vorgehensweise bei der Literaturrecherche, Selektion der wissenschaftlich guten Studien und den Ergebnissen einer großen Anzahl an Patienten wissenschaftlich äußerst wertvoll.**

Leitlinie (engl.: Practice Guideline, Clinical Guideline)

Eine Leitlinie ist eine systematisch entwickelte Empfehlung über eine medizinische und therapeutische Vorgehensweise bei einer bestimmten Krankheit oder Behinderung. Eine gute Leitlinie entspricht dem aktuellen wissenschaftlichen Stand. Sie soll Fachpersonen und Patienten helfen, **Entscheidungen hinsichtlich der Behandlung** zu treffen. Darunter sind nicht nur Medikamente bzw. andere therapeutische Behandlungen zu verstehen, sondern auch z. B. diagnostische Verfahren.

Da eine Leitlinie im klinisch-therapeutischen Alltag direkt umgesetzt werden soll, ist sie sehr praxisorientiert und möglichst den lokalen Gegebenheiten angepasst. Aus diesem Grund gibt es nationale Kompetenzzentren. In Deutschland heißt es **Ärztliches Zentrum für Qualität in der Medizin (ÄZQ)**. Unter anderem ist es für die »Entwicklung und Beurteilung von Methoden und Instrumenten der Qualitätsförderung und Transparenz in der Medizin (z. B. Leitlinien, Qualitätsindikatoren, Patienteninformatio-

nen)« zuständig (Ärztliches Zentrum für Qualität in der Medizin 2008). Das ÄZQ sowie weltweit andere entsprechende Organisationen sind zudem Mitglieder eines internationalen Netzwerkes, dem Guidelines International Network (G-I-N).

Gute Leitlinien befinden sich auf einer hohen Evidenzstufe.

Health Technology Assessment (HTA)

Unter Health Technology Assessment (HTA) versteht man die systematische, evidenzbasierte Bewertung medizinischer Technologien, Prozeduren und Hilfsmittel sowie von Organisationsstrukturen, welche medizinische Leistungen erbringen.

Ein wichtiges Ziel der HTA-Berichte ist, als **Entscheidungshilfe für gesundheitspolitische Fragestellungen** zu dienen, um beispielsweise Innovationen in den Katalog der Leistungsträger im Gesundheitswesen aufzunehmen oder abzulehnen. HTA-Berichte zeichnen sich durch eine breite Sichtweise aus: Über die Gesundheit des Einzelnen hinaus berücksichtigen sie auch die Auswirkungen auf das Gesundheitssystem und die Gesellschaft. Dazu untersuchen sie folgende Aspekte (Deutsches Institut für medizinische Dokumentation und Information 2007):

- experimentelle Wirksamkeit (efficacy)
- Wirksamkeit unter Alltagsbedingungen (effectiveness)
- vergleichende Bewertung der Wirksamkeit (comparative effectiveness)
- gesundheitsökonomische Bewertung (efficiency)
- soziale, rechtliche und ethische Implikationen.

In Deutschland wurde im Jahr 2000 die Deutsche Agentur für HTA (DAHTA) beim Deutschen Institut für medizinische Dokumentation und Information (DIMDI) gegründet. Sie führt ein Programm zur Erstellung von HTA-Berichten durch. Zudem stellt sie eine Datenbank (DAHTA-Datenbank, ► <http://www.dimdi.de/static/de/db/dbinfo/dahta.htm>) und HTA Berichte kostenlos zur Verfügung.

Das Evidenzniveau guter HTAs ist hoch.

2 Gegenüberstellung der qualitativen und quantitativen Forschung

Neben der quantitativen Forschung nimmt die qualitative Forschung eine zunehmend wichtige Position ein. Die qualitative Forschung wird hier oder im ► Buch zwar nicht weiter behandelt, aber eine kurze Gegenüberstellung der beiden unterschiedlichen Forschungsansätze soll den Leserinnen und Lesern einen Eindruck über die wesentlichen Unterschiede verschaffen.

2.1 Quantitative Forschung

Die quantitative Forschung strebt danach, unter möglichst **kontrollierten Bedingungen** meist unter Laborbedingungen zu forschen, **viele Patienten** in die Untersuchung mit einzubeziehen (mit Ausnahme des Single-subject research design und der Fallstudien), **Rückschlüsse auf einzelne Faktoren** zu ziehen und eine Beeinflussung der Untersuchung durch den Experimentator zu vermeiden. Zum Teil wird z. B. im soziologischen oder psychologischen Bereich versucht, auch komplexere Zusammenhänge durch aufwändigere Mehrebenenanalysen zu erfassen, doch diese Untersuchungen bergen aufgrund der Abstraktion die Schwierigkeit der Rückbindung an den untersuchten Alltag (Flick 2000, S. 13–14). Die Auswertungen sind quantitativ, d. h. Beziehungen zueinander (Korrelationen), Unterschiede, prozentuale Anteile etc. werden konkret in Form von Zahlen ausgedrückt und hinsichtlich der statistischen Signifikanz überprüft.

2.2 Qualitative Forschung

Die qualitative Forschung hat dagegen die Ziele, charakteristische Phänomene zu identifizieren, Elemente und deren Beziehungen zueinander aufzudecken und Muster, z. B. Verhaltensmuster oder ein soziales Gefüge, zu bestimmen. Die Wissenschaftler führen die Forschung **nicht unter kontrollierten Bedingungen** durch und zerlegen dabei auch nicht die Untersuchungsgegenstände in einzelne Variablen, sondern sie gestalten sie so offen, dass sie die Gegenstände in ihrer **Komplexität und Ganzheit** in ihrem alltäglichen Kontext untersuchen können. Die qualitative Forschung beschränkt sich meist auf eine **kleine Probendanzahl**, wodurch sich anbietet, ungewöhnliche (selten) Situationen und Personen zu untersuchen. Die Kommunikation des Forschers wird häufig Bestandteil der Erkenntnis. Man betrachtet sie nicht wie bei der quantitativen Forschung als Störvariable und schließt sie nicht möglichst aus. Auch Eindrücke des Forschers sind in der qualitativen Forschung wichtig (Flick 2000, S. 14–16).

Qualitative Studien lassen sich nicht mit den im ► Buch beschriebenen Gütekriterien beurteilen. Aufgrund ihrer spezifischen Ziele und Methoden erfordern sie **eigene Beurteilungskriterien**. Hierfür sei auf weitere Literatur verwiesen, insbesondere auf die Artikel von Giacomini u. Cook (2000a, 2000b) sowie auf das Kriterienformular von Letts et al. (2007).

3 Forschungsrichtungen

Für die verschiedenen Themenbereiche bieten die EBM und EBP Tabellen, welche viele der oben beschriebenen Studiendesigns hinsichtlich ihrer Evidenzstufe einordnen (als Beispiel ► Buch, Tab. 18.1). Darin tauchen jedoch nicht nur Studiendesigns auf, sondern es fallen auch Begriffe über Forschungsrichtungen wie Bench research. Diese werden nun kurz erklärt.

3.1 Klinische Forschung

Die klinische Forschung führt Studien an Patienten durch (► Abschn. 1.1). Sie findet unter kontrollierten Bedingungen statt, indem die Wissenschaftler z. B. Einschlusskriterien und das experimentelle Vorgehen vorgeben und die Störvariablen weitgehend reduzieren. Sie schaffen Evidenz unter kontrollierten und häufig idealtypischen Bedingungen.

3.2 Bench Research (biomedizinische Grundlagenforschung)

Bench Research findet im Forschungslabor statt und beinhaltet keine Versuche am Menschen, sondern Experimente mit **Zellen, Zellkulturen, Gewebeprobe**n und **Tiermodellen**. Ziel ist herauszufinden, wie spezifische Gene bzw. Moleküle oder zelluläre Mechanismen mit physiologischen oder pathologischen Systemen verbunden sind, um letztendlich bessere Möglichkeiten zur Krankheitsbekämpfung zu finden.

3.3 Outcomes Research (Versorgungsforschung)

Die Versorgungsforschung beschreibt, analysiert und interpretiert die gesundheitliche **Versorgung** Einzelner oder von Teilen der Bevölkerung mit gesundheitsrelevanten Produkten und Dienstleistungen **unter Alltagsbedingungen**. Dazu gehören

Untersuchungen des Versorgungsbedarfs, der Versorgungsstrukturen bzw. -prozesse, der erbrachten Versorgungsleistungen und des Zugewinns an Gesundheits- bzw. Lebensqualität.

Neben herkömmlichen Studiendesigns wie die Kohortenstudie gibt es spezifische Studienarten in der Versorgungsforschung:

- **Anwendungsbeobachtung:** Die Anwendungsbeobachtung sammelt mit Hilfe von Erhebungsbögen Daten, um z. B. den Nutzen von Behandlungsmaßnahmen in der breiten routinemäßigen Anwendung zu bestimmen. Eine Vergleichsgruppe fehlt, die Studie ist deskriptiv. Die Anwendungsbeobachtung ist besonders geeignet, um Fragestellungen bezüglich Prognose, Überlebenszeit, Lebensqualität, Nutzen und Therapiekosten zu beantworten.
- **Post-authorisation Safety Study (PASS):** Die PASS ist eine Unterart der Anwendungsbeobachtung und beschäftigt sich mit der Sicherheit therapeutischer Interventionen: Sie dient der Ermittlung der Häufigkeit bereits bekannter, unerwünschter Wirkungen therapeutischer Interventionen unter Praxisbedingungen und der Identifizierung bisher unbekannter, seltener Sicherheitsprobleme, welche klinische Studien aufgrund begrenzter Patientenzahlen nicht aufdeckten.

Die Versorgungsforschung bedient sich zudem verschiedener klinischer Studiendesigns, um viele Informationen zusammenzutragen. Darunter zählen insbesondere:

- Kohortenstudien,
- Fall-Kontroll-Studien,
- gesundheitsökonomische Studien,
- Health Technology Assessments (HTA),
- kontrollierte klinische Studien,
- systematische Reviews,
- Meta-Analysen.

Praxistipp

Für viele Fragestellungen mangelt es zurzeit noch an der Anzahl und wissenschaftlichen Qualität von Forschungsarbeiten bei den Medizinalfachberufen. Trotzdem lohnt sich die Suche nach der **besten zur Verfügung stehenden Evidenz**.

Ergebnisse von Forschungsarbeiten, deren wissenschaftliche Qualität sich auf niedrigem Niveau befindet, sollten Sie relativieren, indem Sie sie mit Vorbehalt akzeptieren, im Sinne von: »Es könnte sein, dass...« oder »Es gibt gewisse Anhaltspunkte dafür, dass...«.

Wenn Ihre Fragestellung für Ihre praktische Arbeit auch 2 Jahre später noch wichtig ist, sollten Sie die Suche nach wissenschaftlicher Evidenz erneut durchführen, denn dann gibt es vielleicht bereits neue und/oder wissenschaftlich fundiertere Erkenntnisse.

Literatur

- Ärztliches Zentrum für Qualität in der Medizin (2008) Tätigkeitsbericht 2007. <http://www.aezq.de/edocs/pdf/taetigkeitsberichte/tb07.pdf>. Zugriffen 15. Aug. 2008
- Atabaki SM, Stiell IG, Bazarian JJ, Sadow KE, Vu TT, Camarca MA, Berns S, Chamberlain JM (2008) A clinical decision rule for cranial computed tomography in minor pediatric head trauma. *Arch Pediatr Adolesc Med* 162(5):439-445
- Blakely TA, Woodward AJ (2000) Ecological effects in multi-level studies. *J Epidemiol Community Health* 54:367-374
- Deeks JJ (2002) Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 21(11):1575-1600
- Deutsches Institut für medizinische Dokumentation und Information (2007) HTA beim DIMDI. http://www.dimdi.de/static/de/hta/basisinfo_de.pdf. Zugriffen 16. Aug. 2008
- Donner-Banzhoff N (2000) Autorenmanual »Levels of Evidence«. <http://www.degam.de/leitlinien/evidence.html>. Zugriffen 21. März 2009
- Ekkernkamp M, Lühmann D, Raspe H (2003) Methodenmanual für »HTA-Schnellverfahren« UND Exemplarisches »Kurz-HTA«: Die Rolle der quantitativen Ultraschallverfahren zur Ermittlung des Risikos für osteoporotische Frakturen. In: Schwartz FW, Gerhardus A, Köbberling J, Raspe H, Graf von der Schulenburg J-M (Reihenherausgeber). *Health Technology Assessment – Schriftenreihe des Deutschen Instituts für Medizinische Dokumentation und Information im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung*, Bd. 34. Asgard-Verlag, Sankt Augustin
- Flick U. Qualitative Forschung (2000) Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. Rowohlt Enzyklopädie, Rowohlt Taschenbuch Verlag, Reinbeck bei Hamburg
- Giacomini MK, Cook DJ for the Evidence-Based Medicine Working Group (2000a) Users' guides to the medical literature: XXIII. Qualitative research in health care A. Are the results of the study valid? *JAMA* 284(3):357-362
- Giacomini MK, Cook DJ for the Evidence-Based Medicine Working Group (2000b) Users' guides to the medical literature: XXIII. Qualitative research in health care B. What are the results and how do they help me care for my patients? *JAMA* 284(4):478-482
- Gitlin LN, Corcoran M, Winter L, Boyce A, Hauck WW (2001) A randomized, controlled trial of a home environmental intervention: effect on efficacy and upset in caregivers and on daily function of persons with dementia. *Gerontologist* 41(1):4-14
- Haneuse S, Wakefield J (2008) Geographic-based ecological correlation studies using supplemental case-control data. *Stat Med* 27(6):864-887
- Laupacis A, Sekar N, Stiell IG (1997) Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA* 277:488-494
- Letts L, Wilkins S, Law M, Stewart D, Bosch J, Westmorland M (2007) Critical Review Form – Qualitative Studies (Version 2.0). http://www.srs-mcmaster.ca/Portals/20/pdf/ebp/qualreview_version2.0.pdf. Zugriffen 21. März 2009
- Lipsey MW, Wilson DB (2001) *Practical Meta-Analysis*. Applied Social Research Methods, Vol. 49. Sage Publications, Inc, Thousand Oaks, London, New Delhi
- MEDDEV 2.7.1, Rev 3 (2009) Guidelines on Medical Devices, Clinical Evaluation: A Guide for Manufacturers and Notified Bodies. European Commission Enterprise and Industry Directorate General. http://ec.europa.eu/consumers/sectors/medical-devices/files/meddev2_7_1rev_3_en.pdf. Zugriffen 09. Juni 2013
- Oxford Centre for Evidence-based Medicine (2009) Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009). <http://www.2cch.org.tw/ebm/file/CEBM-Levels-of-Evidence.pdf>. Zugriffen 17. Dez. 2010
- Stiell IG, Wells GA (1999) Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 33(4):437-447
- Turner-Stokes L (1999) Introduction: Effectiveness of rehabilitation. *Clin Rehabil* 13(suppl 1):3-6

Kritische Bewertung gesundheitsökonomischer Studien

Einfach ist die Wahl zwischen zwei Interventionen, wenn beide gleich viel kosten und eine davon dem Patienten einen höheren Nutzen bringt. Was aber, wenn die wirksamere teurer ist? Für eine Art des Nutzens gibt es Empfehlungen, wie viel höher dieser sein sollte, um die höheren Kosten der wirksameren Intervention zu rechtfertigen. »Eine Art des Nutzens« deutet an, dass es mehrere gibt – es ist interessant, wie vielfältig »Nutzen« in der Gesundheitsökonomie definiert wird und wie ökonomische Studien diesen Nutzen und die Kosten verschiedener Interventionen miteinander vergleichen.

1 Notwendigkeit und Ziel gesundheitsökonomischer Studien

In den letzten Jahren wurden die Kosten bzw. die Kostenexplosion im Gesundheitswesen ein immer wichtigeres Thema. Therapeutinnen und Therapeuten erfahren in ihrem Praxisalltag, dass **finanzielle Ressourcen** sowohl bei Kostenträgerinnen als auch im klinischen Umfeld häufig sehr beschränkt sind, was u. a. zu geringer personeller Besetzung und damit knapp bemessener Therapiezeit pro Patient, unbefriedigender materieller Ausstattung der therapeutischen Umgebung, qualitativ oder quantitativ unzureichender Hilfsmittelversorgung, mangelnden Möglichkeiten der altersgerechten Unterbringung des Patienten in geeigneten Wohnheimen etc. führen kann. Daher gewinnen gesundheitsökonomische Studien zunehmend an Bedeutung. Sie sollen helfen, Interventionen zu selektieren, die uns den »biggest bang for the buck« (Walton et al. 1994), frei übersetzt den »größten Knall für die Knete«, bringen. Dabei sind mit Interventionen nicht nur Handlungen am erkrankten Patienten gemeint, sondern auch prophylaktische Maßnahmen (z. B. Reihenuntersuchungen in Schulen, Fitness-Programme für ältere Menschen), Einrichtungen spezieller Institutionen (z. B. Wohnheime, therapeutische Zentren z. B. in Sonderschulen, Selbsthilfegruppen) etc.

➤ **Ökonomische Analysen der Kosten und Nutzen von Interventionen dienen als Entscheidungshilfe, welche Intervention aus alternativen Vorschlägen zu bevorzugen ist. Sie sind nicht nur für Therapeuten und weitere Fachleute des Gesundheitswesens wichtig, sondern auch z. B. für Gesetzgeber und Kostenträger wie Krankenkassen und Unfallversicherungen.**

2 Analyse gesundheitsökonomischer Studien in der EBP

Ökonomische Studien lassen sich mit denselben übergeordneten Fragen analysieren, welche auch in den anderen Themenbereichen der EBP (► Buch; Mangold: Evidenzbasiertes Arbeiten in der Physio- und Ergotherapie: Kap. 8-20) zu finden sind: die Fragen nach der Validität, Bedeutsamkeit und Anwendbarkeit der Evidenz.

■ Die Analyse der **Validität** einer ökonomischen Studie zeigt, ob die Studie wirklich bestimmt, welche der untersuchten Interventionen den besten Nutzen für die vorhandenen

ökonomischen Ressourcen bringt. Wie stark die Evidenz ist, hängt von der Einhaltung der internen Gütekriterien ab. Zu ihrer Bewertung sind weiter unten Leitfragen aufgeführt.

- Die **Bedeutsamkeit** der Evidenz beurteilt man anhand der in der Studie durchgeführten ökonomischen Berechnungen. Wichtig sind dabei nicht nur die ökonomischen Unterschiede als punktuelle Werte, die praktisch nie ganz genau sind, sondern auch deren Bandbreiten aufgrund möglicher Fehlerquellen in der Datenerfassung.
- Die Frage nach der **Anwendbarkeit** analysiert, ob sich die Resultate der Studie auf die eigene klinisch-therapeutische Situation bzw. auf den eigenen Patienten übertragen lassen.

Ökonomische Studien sind bisher seltener als andere Studien im Gesundheitswesen anzutreffen. Das zugehörige Fachvokabular, die verschiedenen Arten ökonomischer Studien und deren Berechnungen sind daher nicht nur vielen Praktikern, sondern auch Wissenschaftlern wenig oder gar nicht vertraut. Deshalb liefern die nachfolgenden Abschnitte zunächst **Grundlagen** über die grundsätzlichen Inhalte dieser Studien, klären, was unter Kosten und Nutzen zu verstehen ist und stellen 3 wichtige Formen gesundheitsökonomischer Analysen vor. Erst danach folgen die Erläuterungen zu den 3 übergeordneten Fragen hinsichtlich der Validität, Bedeutsamkeit und Anwendbarkeit der Evidenz.

Es sei an dieser Stelle darauf hingewiesen, dass die Gesundheitsökonomie ein komplexes Gebiet ist. Hier werden Begriffe und Vorgehensweisen nur grob erklärt. Für eine vertiefte Auseinandersetzung mit dem Thema ist es unumgänglich, spezifische gesundheitsökonomische Bücher hinzuzuziehen, z. B. Brazier et al. (2007) oder Drummond et al. (2005).

3 Grundprinzip ökonomischer Studien

Ökonomische Studien im engeren Sinne berechnen nicht einfach nur die **Kosten** der Interventionen, sondern sie berücksichtigen auch den **Nutzen**. Wichtig ist ebenso, dass sie einen **Vergleich** zwischen der zu bewertenden Intervention und einer oder mehreren Vergleichsalternativen ziehen, z. B. den Vergleich mit

- keiner Behandlung bzw. einer Placebobehandlung,
- einer herkömmlichen Behandlung, d. h. mit einer Behandlung, die sehr häufig bzw. am häufigsten in der Praxis angewendet wird (Standardintervention, konventionelle Intervention),
- einer Behandlung, die therapeutisch als optimal gilt (Goldstandard). Dieser Goldstandard kann, muss aber nicht identisch mit der Standardintervention sein,
- weiteren neuen Behandlungen.

➤ **Das Grundprinzip ökonomischer Studien »besteht im Vergleich von Ressourceneinsatz (Input) und Gesundheitsverbesserung (Output)«, Klier u. Pogantsch 2007.**

Den **Input** stellen die aufgewendeten Kosten dar. Der untersuchte **Output** richtet sich nach der Art der ökonomischen Analyse, z. B.:

- ein Faktor, welcher die Änderung des Gesundheitszustandes beschreibt, z. B. die Verbesserung der Selbständigkeit im Alltag,
- die Anzahl gewonnener Lebensjahre mit maximaler Lebensqualität,
- monetäre Einheiten, also Geld, z. B. durch die wiedergewonnene Arbeitsleistung eines Patienten.

Diese 3 wichtigsten Formen des Outputs sind weiter unten genauer beschrieben, ► Abschn. 6.1, ► Abschn. 6.2, ► Abschn. 6.3.

4 Definition und Arten der Kosten

In der Ökonomie basieren die Kosten auf dem Wert, den man gewinnen würde, wählte man eine andere Möglichkeit. Oder anders ausgedrückt: Führt man eine Intervention durch, so stehen die darauf verwendeten finanziellen Ressourcen nicht mehr anderen Interventionen zur Verfügung, und deren möglicher Nutzen wurde geopfert (Phillips u. Thompson 2001a). Die Kosten heißen daher auch **Opportunitätskosten**.

Folgende verschiedenen Arten von Kosten werden unterschieden (Phillips u. Thompson 2001a):

1. **Direkte Kosten:** z. B. Medikamente, Krankenhauskosten, Kosten für medizinisch-therapeutisches Personal, Transportkosten für den Patienten etc.
2. **Indirekte Kosten:** z. B. Produktivitätsverluste durch (vorübergehende) Arbeitsunfähigkeit des Patienten oder medizinische Geräte, die in der Zeit nicht für andere Zwecke zur Verfügung stehen.
3. **Immaterielle Werte** (engl.: intangibles): z. B. Schmerzen, Nebenwirkungen.

Es ist in der Praxis üblich anzunehmen, dass der Preis, welcher tatsächlich gezahlt wurde, die Opportunitätskosten wiedergibt (Phillips u. Thompson 2001a).

5 Gesundheitlicher Nutzen

Der gesundheitliche Nutzen wird sehr unterschiedlich ausgedrückt, beispielsweise als Anzahl geretteter Leben oder gewonnener Lebensjahre, als Grad der Arbeitsfähigkeit des Patienten oder als Geldbetrag z. B. durch vermiedene Arbeitsunfähigkeit.

Aufgrund der verschiedenen Arten des gesundheitlichen Nutzens lassen sich **mehrere Formen gesundheitsökonomischer Studien** unterscheiden. Die nächsten Abschnitte beschreiben sie kurz, um einen Eindruck zu vermitteln, worum es bei den Studien geht bzw. wie sie sich unterscheiden. Weiter unten werden die darin enthaltenen Analysen genauer erklärt.

6 Wichtige Formen gesundheitsökonomischer Studien

Wichtige 3 Formen der gesundheitsökonomischen Studien sind die:

- Kosten-Effektivitäts-Analyse (Cost-Effectiveness Analysis)
- Kosten-Nutzwert-Analyse (Cost-Utility Analysis)
- Kosten-Nutzen-Analyse (Cost-Benefit Analysis)

6.1 Kosten-Effektivitäts-Analyse

Die Kosten-Effektivitäts-Analyse, auch Kosten-Wirksamkeits-Analyse oder im Englischen Cost-Effectiveness Analysis (CEA) genannt, drückt den gesundheitlichen Nutzen auf **nicht monetärer** (nicht geldlicher) Ebene aus, beispielsweise durch die Anzahl gewonnener Lebensjahre, symptomfreier Tage, Anzahl bzw. Anteil Patienten mit dem gewünschten gesundheitlichen Effekt oder bei Messinstrumenten auch z. B. durch die Anzahl entdeckter Fälle einer bestimmten Krankheit. Wichtig ist, dass sich der gesundheitliche Nutzen immer auf einen **Vergleich zwischen mehreren Möglichkeiten** bezieht, also beispielsweise zwischen 2 Behandlungen oder zwischen einer Behandlung und keiner Behandlung.

Beispiel Kosteneffektivität

Die Studie von Luijsterburg et al. (2007) untersuchte die Kosteneffektivität einer physiotherapeutischen Behandlung, welche zusätzlich zur generellen Versorgung durch den Hausarzt bei Patienten mit einer Ischialgie durchgeführt wurde, im Vergleich zur alleinigen Versorgung durch den Hausarzt.

Eine der Zielgrößen war der »globale wahrgenommene Effekt« auf einer zunächst 7-stufigen Skala (1= vollständige Heilung, ..., 7= Verschlechterung in hohem Maße), welche zur Auswertung in »verbessert« (mit mindestens hoher Verbesserung) und »nicht verbessert« (mit höchstens geringer Verbesserung) dichotomisiert wurde. Der Nutzen war hier daher die Anzahl bzw. der Anteil Patienten mit dem gewünschten gesundheitlichen Effekt »Verbesserung«.

Die Studie kam zum Ergebnis, dass das inkrementelle Kosten-Effektivitäts-Verhältnis (ICER, ► Abschn. 9.1, »Inkrementelles Kosten-Effektivitäts-Verhältnis«) nach 1 Jahr 6224 € pro zusätzlichem Patient mit positivem Ergebnis betrug. Aufgrund dieser Höhe nahmen die Autoren der Studie an, dass die Physiotherapie keine kosteneffektive Zusatzbehandlung bei dieser Indikation darstellt.

Der gesundheitliche Nutzen muss für die Kosten-Effektivitäts-Analyse noch mit den Kosten verrechnet werden. Wie das geschieht, wird weiter unten bei der Frage erklärt, ob die Analyse ein bedeutsames Ergebnis zeigt.

Der Nachteil der Kosten-Effektivitäts-Analyse ist, dass sie keine Interventionen vergleichen kann, die sich in mehr als einem Parameter unterscheiden (Brazier et al. 2007, S. 10). Für ein **komplexeres Konstrukt** aus mehreren Parametern, aus welchem sich die Gesundheit zusammensetzt, sind andere Methoden notwendig, z. B. die nachfolgend beschriebene Kosten-Nutzwert-Analyse.

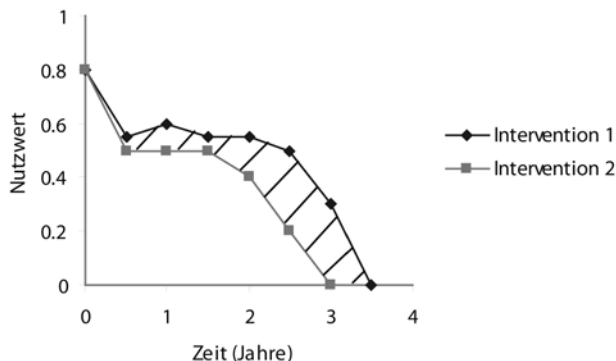
6.2 Kosten-Nutzwert-Analyse

Die Kosten-Nutzwert-Analyse (Cost-Utility Analysis, CUA) ist eine spezielle Variante der Kosten-Effektivitäts-Analyse. Um den Nutzen auszudrücken, kombiniert sie eine quantitative Komponente, meist die Anzahl gewonnener Lebensjahre, mit einer qualitativen Komponente, meist eine Bewertung der Lebensqualität. Obwohl es verschiedene Möglichkeiten für die Art des Nutzwertes gäbe, hat sich durchgesetzt, ihn in Form der qualitätsangepassten Lebensjahre (engl.: **Quality-adjusted-life-years, QALYs**) auszudrücken. Diese Größe – der nachfolgende Text verwendet das gebräuchliche Akronym QALYs – bedarf einer genaueren Erklärung.

■ **Tab. 1** Beispiele verschiedener Gesundheitsprofile, erfasst mit dem EQ-5D

Patient	Mobilität	Selbstversorgung	Übliche Aktivitäten	Schmerzen/Unwohlsein	Ängstlichkeit/Depression
Meier	1	1	1	1	1
Krause	2	1	3	2	1
Steger	3	3	3	1	3
Moser	2	3	1	2	3
...

Die Zahlen bedeuten Stufen: 1= jeweils schlechterer Zustand, 3= bester Zustand in der betreffenden Dimension



■ **Abb. 1** Quality-adjusted-life-years. Die gestrichelte Fläche entspricht der Anzahl QALYs, welche die Patienten mit der Intervention 1 im Vergleich zur Intervention 2 gewinnen würden

Quality-adjusted-life-years (QALYs)

Wie oben bei der Kosten-Effektivitäts-Analyse erwähnt, lässt sich der Nutzen einer Behandlung in Form **gewonnener Lebensjahre** ausdrücken. Allerdings sagt diese Größe nichts darüber aus, in welchem Maß diese Lebensjahre **lebenswert** sind. Die Anzahl gewonnener Lebensjahre gibt zwar Auskunft über die Quantität, nicht aber über die Qualität. Um diese beiden Komponenten zu verbinden, wurde die Größe QALYs (engl.: Quality-adjusted-life-years), die qualitätsangepassten Lebensjahre, eingeführt. 1 QALY entspricht 1 zusätzlichen Lebensjahr mit perfekter Gesundheit, welches der Patient durch eine Intervention gewinnt. Wenn er beispielsweise 2 zusätzliche Jahre mit »halbperfekter« Gesundheit oder 3 Jahre mit »einem Drittel perfekter« Gesundheit gewinnt, so entsprechen diese Beispiele jeweils 1 QALY. Der Vergleich zweier Interventionen hinsichtlich der QALYs lässt sich graphisch darstellen (■ Abb. 1).

Die Anzahl QALYs geben die Studien direkt an. Die Leser müssen sie also nicht selbst bestimmen. Für das Verständnis der Kosten-Nutzwert-Analysen ist es jedoch gut zu wissen, wie die Wissenschaftler und Wissenschaftlerinnen ermitteln, was eine perfekte oder z. B. »halbperfekte« Gesundheit ist und wie sie die QALYs berechnen. Deshalb wird der Vorgang hier grob erklärt.

Zur Berechnung der QALYs sind 2 Schritte nötig, nämlich die **Beschreibung** und die **Bewertung** des Gesundheitszustandes bzw. Gesundheitsprofils des Patienten. Beide Schritte zusammen heißen auch **präferenzbasierte Messung der Gesundheit** (engl.: preference-based measure of health, Brazier et al. 2007, S. 11). Was beinhalten diese 2 Schritte nun?

1. Schritt: Beschreibung des Gesundheitszustandes bzw. -profils

Der Gesundheitszustand wird durch spezifische oder generische (umfassendere, ganzheitliche) Erfassungsinstrumente im Verlauf der Zeit immer wieder bestimmt, beispielsweise in Abständen von einem halben Jahr.

- Zu den **spezifischen Instrumenten** gehören z. B. solche zur Erfassung von Rückenschmerzen. Sie bieten u. a. den Vorteil, dass sie den betreffenden Bereich besser als generische Erfassungen abdecken.
- **Generische Erfassungsinstrumente** weisen den Vorteil auf, dass sie sich über viele Bereiche erstrecken, welche die Gesundheit mitbestimmen. Auch sie setzen jeweils bestimmte Schwerpunkte, daher erfassen verschiedene generische Instrumente unterschiedliche Daten.

Beispiel Generisches Erfassungsinstrument

Das **EuroQol-Klassifikationssystem (EQ-5D)** (The EuroQol Group 1990) ist ein häufig gebrauchtes generisches Erfassungsinstrument für ökonomische Analysen. Es erfasst 5 Dimensionen: Mobilität, Selbstversorgung, übliche Aktivitäten (Arbeit, Ausbildung, Freizeitaktivitäten), Schmerzen/Unwohlsein und Ängstlichkeit/Depression. Die Beurteilung dieser 5 Dimensionen erfolgt je mithilfe dreier Stufen: Bei der Dimension Schmerz beispielsweise bedeutet die 1. Stufe »kein Schmerz oder Unwohlsein«, die 2. Stufe »moderater Schmerz oder Unwohlsein«, die 3. Stufe »extremer Schmerz oder Unwohlsein«. Das Gesundheitsprofil jedes Patienten setzt sich damit aus 5 verschiedenen Zahlen zusammen (■ Tab. 1).

Aus der Tabelle geht hervor, dass sich sehr viele Gesundheitsprofile ergeben: bei Frau Meier 11111, Herrn Krause 21321 etc. Insgesamt gibt es, inklusive »bewusstlos« und »tot«, **245 mögliche Gesundheitsprofile**.

Zur Auswertung dieser Gesundheitsprofile darf man nicht einfach die Punkte zusammenzählen, um zur Lebensqualität zu gelangen. Falsch ist also: Frau Meier hat 5 Punkte, Herr Krause 9 Punkte etc. Vielmehr beinhaltet jedes der 245 Profile einen eigenen Lebensqualitätswert, welcher wissenschaftlich ermittelt wurde (► 2. Schritt).

2. Schritt: Bewertung des Gesundheitszustands bzw. -profils

Wie vorhergehend beschrieben, beinhalten Instrumente wie das EQ-5D Hunderte möglicher Gesundheitszustände bzw. -profile. Für ökonomische Analysen ist es notwendig zu wissen, welches Profil welchen **Lebensqualitätswert** hat. Dazu gibt es Studien, welche die Werte für alle Gesundheitsprofile ermitteln. Sie befragen dazu Menschen aus der generellen Bevölkerung oder Betroffene einer Krankheit, um herauszufinden, wie sie die Profile einschätzen. Bei der Befragung werden wissenschaftliche Methoden angewendet, v. a. das **standard gamble (SG)**, **time trade-off (TTO)** und **visual analog scale (VAS)**.

Es würde den Rahmen dieser Publikation sprengen, die Vorgehensweisen zu erklären, zumal diese 3 Formen Varianten aufweisen. Es ist wichtig zu wissen, dass das Ergebnis der ökonomischen Analyse davon beeinflusst wird, welche dieser 3 Formen und welche zugehörige Variante für die Bewertung der Lebensqualität gebraucht wurde. Wenn 2 ökonomische Artikel unterschiedliche Verfahrensweisen benutzt haben, sind sie **nicht direkt miteinander vergleichbar**, es sei denn, es werden

■ **Tab. 2** Beispiele einer Liste zur Konvertierung der EQ-5D Gesundheitsprofile in die Lebensqualitätswerte (Phillips u. Thompson 2001b)

Gesundheitszustand	Beschreibung	Lebensqualitätswerte
11111	Keine Probleme	1,0
12321	Ein paar Probleme beim Gehen, ein paar Probleme bei der Körperpflege oder dem Anziehen, unselbstständig bei der Verrichtung üblicher Aktivitäten (Arbeit, Ausbildung, Freizeitaktivitäten), moderater Schmerz oder Unwohlsein, keine Ängstlichkeit oder Depression	0,329
23322	Ein paar Probleme beim Gehen, unselbstständig bei der Körperpflege oder beim Anziehen, unselbstständig bei der Verrichtung üblicher Aktivitäten (Arbeit, Ausbildung, Freizeitaktivitäten), moderater Schmerz oder Unwohlsein, moderate Ängstlichkeit oder Depression	0,079

geeignete Verfahren verwendet, welche solche Verzerrungen kompensieren (Brazier et al. 2007, S. 107–109).

Die Anwendung der SG, TTO und VAS führt letztendlich zu einer Liste der möglichen Gesundheitszustände bzw. -profile mit ihren zugehörigen Bewertungen der Lebensqualität, welche eine Spanne von **–Unendlich bis 1** umfassen:

- 1 bedeutet eine perfekte Gesundheit.
- 0 bedeutet, dass die Lebensqualität gleichwertig mit dem Tod ist.
- Eine negative Zahl zeigt einen Zustand schlimmer als den Tod an.

Wenn eine Studie beispielsweise den Gesundheitszustand von Menschen mit einer Tetraplegie beschreibt, so muss sie deren erfasste Gesundheitsprofile hinsichtlich der Lebensqualität bewerten. Dies geschieht mithilfe von Algorithmen, die Wissenschaftler für das Instrument spezifisch entwickelt haben (■ Tab. 2, Beispiele bereits errechneter Lebensqualitätswerte für das Erfassungsinstrument EQ-5D). Die Umrechnungen, welche dieses Instrument beinhalten, beschrieb Dolan (1997).

Wissenschaftler gebrauchen diese Liste (■ Tab. 2), um die ermittelten Gesundheitsprofile der Patienten in Lebensqualitätswerte zu konvertieren, um damit die Anzahl QALYs zu ermitteln.

- **Die Anzahl QALYs ist das Produkt aus den zusätzlich gewonnenen Lebensjahren und dem Lebensqualitätswert dieser Lebensjahre.**

Für die Berechnung der QALYs einer Patientengruppe ist allerdings zu berücksichtigen, dass über einen längeren Zeitraum verschiedene Gesundheitszustände auftreten können. Deshalb gewichtet man **jeden Zeitabschnitt** eines bestimmten Gesundheitszustandes mit seinem Lebensqualitätswert. Die **Gewichtung** besteht aus einer Multiplikation der Zeitspanne mit dem Lebensqualitätswert (► Beispiel QALY-Berechnung). Am Schluss werden alle gewichteten Zeitabschnitte addiert.

- **1 QALY entspricht 1 Lebensjahr mit perfekter Gesundheit.**

Beispiel QALY-Berechnung

Ein Patient lebt nach einer Behandlung noch 5 Jahre. In den ersten 2 Jahren beträgt der Lebensqualitätswert 0,8, in weiteren 2 Jahren 0,6 und im letzten Lebensjahr 0,2.

Daraus ergibt sich insgesamt folgender Nutzwert:

$$2 \times 0,8 + 2 \times 0,6 + 1 \times 0,2 = 3 \text{ QALYs.}$$

Wie ist der Wert 3 QALYs zu interpretieren? 3 QALYs bedeuten 3 Jahre mit einem perfekten Gesundheitszustand. Dieser Wert ist eine **Vereinfachung**, denn der Patient im Beispiel lebte ja länger und hatte die ganze Zeit über keinen perfekten Gesundheitszustand. Aber diese Zahl erhält dann ihren Sinn, wenn man den Nutzen zweier Behandlungen – oder einer Behandlung im Vergleich zu keiner Behandlung – miteinander vergleicht.

Auswirkungen unterschiedlicher Vorgehensweisen auf die Ergebnisse

Die unterschiedlichen Arten zur Beschreibung und Bewertung des Gesundheitszustandes bzw. -profils wirken sich auf die Ergebnisse, d. h. auf die Anzahl QALYs, aus. Daher sind Studien, welche verschiedene Vorgehensweisen zur Erfassung und Bewertung gewählt haben, **nicht direkt vergleichbar**. Beispielsweise zeigte eine Studie, dass das QWB¹, 15D² und SF-6D³ höhere Mittelwerte ergaben als das EQ-5D (EuroQol-5 Dimensions), HUIs⁴ und AQoL⁵ (Hawthorne et al 2001).

Weitere wichtige Faktoren, welche die Berechnung der QALYs bzw. Kosten-Nutzwert-Analysen beeinflussen und daher zur mangelnden Vergleichbarkeit führen können, sind z. B. die Technik der Bewertung (SG, TTO, VAS) und deren Varianten (► 1. Schritt), die Parameter bei der Zusammenfassung der Daten der verschiedenen Befragten (z. B. Mittelwert, Median, Modalwerte) und die Gruppe, welche die Gesundheitszustände bewertet.

Die QALYs müssen für die Kosten-Nutzwert-Analyse noch mit den **Kosten** verrechnet werden. Wie das geschieht, wird weiter unten bei der Frage erklärt, ob das Ergebnis praxisrelevant ist (► Abschn. 9.2).

6.3 Kosten-Nutzen-Analyse (Cost-Benefit Analysis, CBA)

Kosten und Nutzen

In der Kosten-Nutzen-Analyse werden nicht nur die Kosten, sondern auch der gesundheitliche Nutzen in **monetärer** Form ausgedrückt:

- Die **Kosten** setzen sich aus den direkten, indirekten und immateriellen Kosten, sofern sich diese bestimmen lassen, zusammen.
- Der **Nutzen** lässt sich ebenfalls in einen direkten und indirekten Nutzen aufteilen (Klier u. Pogantsch 2007):
 - Der **direkte** Nutzen entsteht durch Einsparungen an medizinischen Dienstleistungen, Medikamenten etc., durch die Verbesserung des Gesundheitszustandes oder Verminderung des Fortschreitens der Krankheit.

- 1 QWB: Quality of Well-Being scale (Kaplan u. Anderson 1988).
- 2 15D: Fifteen-dimensional measure of health related quality of life (Sintonen u. Pekurinen 1993; Sintonen 1994).
- 3 SF-6D (Short-Form-6 Dimensions): Ein Abkömmling des SF-36 und SF-12 (Brazier et al. 2002; Brazier et al. 2004).
- 4 HUI: Health Utility Index. Es werden verschiedene unterschieden: HUI1, HUI2 und HUI3 (Torrance 1982; Torrance et al. 1996; Feeny et al. 2002).
- 5 AQoL: Australian Quality of Life instrument (Hawthorne et al. 1997).

- Der **indirekte** Nutzen kommt beispielsweise durch die Verminderung von Produktivitätsverlusten, durch vermiedene Arbeitsunfähigkeit oder verhin- derte Frühpensionierung zustande.

➤ **Die Kosten-Nutzen-Analyse ermittelt das Verhältnis von Kosten und Nutzen in monetärer Form.**

Vor- und Nachteile der Kosten-Nutzen-Analyse

Der **Vorteil** der Kosten-Nutzen-Analyse besteht in der gleichen Einheit der Kosten und des Nutzens. Durch die Berechnung des medizinisch-therapeutischen Nettonutzens, d.h. der Differenz zwischen Kosten und Nutzen, ist es sogar möglich, das Projekt aus dem Gesundheitsbereich mit Projekten aus ganz anderen Bereichen zu vergleichen, z.B. mit erhöhten Anstrengungen im Bildungswesen (Walshe u. Diehl 1998).

Der **Nachteil** der Kosten-Nutzen-Analyse liegt darin, dass »keineswegs sämtlicher Nutzen erfasst und monetär bewertet werden kann; dies beginnt beispielsweise beim Wert des menschlichen Lebens« (Klier u. Pogantsch 2007). Diese Aussage führt zu der Frage: Wie wird der monetäre Nutzen überhaupt bestimmt?

Bestimmung des monetären Nutzens

Zur Bestimmung des monetären Nutzens existieren 2 Verfahren: das kostentheoretische Verfahren und der Zahlungsbereitschaftsansatz. Sie lassen sich folgendermaßen charakterisieren (Walshe u. Diehl 1998):

- **Kostentheoretischer Ansatz:** Bei diesem Verfahren besteht der direkte monetäre Nutzen in den Behandlungskosten, welche eine medizinische bzw. therapeutische Behandlung vermeidet, z.B. die Einsparung von Krankenhauskosten durch eine neue ambulante Versorgungsmaßnahme. Der indirekte Nutzen besteht in der Vermeidung des Einkommensverlustes, welcher ohne die Intervention bzw. trotz einer Standardintervention eintreffen würde. Hierin liegt eine Schwäche, denn die Höhe des Einkommens beeinflusst die Kosten.
- **Zahlungsbereitschaftsansatz:** Dieser Ansatz ermittelt den Nutzen anhand des Geldbetrages, welches die Patienten für die Behandlung mit deren entsprechendem Erfolg zu zahlen bereit wären. Die Beträge erfassen die Wissenschaftler mittels Befragungen oder Verhaltensbeobachtungen. Das Ergebnis ist aber mit einem starken Unsicherheitsfaktor behaftet.

Nachdem geklärt ist, welche Vergleiche die 3 wichtigen Formen der gesundheitsökonomischen Studien zwischen Kosten und Nutzen ziehen, stellt sich die Frage, wie solche Studien in wissenschaftlicher Hinsicht konzipiert sind. Handelt es sich beispielsweise auch um randomisiert kontrollierte Studien, wie sie z.B. bei Wirksamkeitsstudien (► Buch; Mangold: Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie: Kap. 18) vorgekommen sind, oder sind die Studiendesigns vollkommen anders?

7 Studiendesigns und -inhalte

Gesundheitsökonomische Studien untersuchen, ob sich eine Intervention aus ökonomischer Sicht lohnt bzw. welche Intervention **aus alternativen Vorschlägen zu bevorzugen** ist. Häufig handelt es sich dabei um Studien, die herausfinden möchten, ob eine Therapie wie z.B. ein therapeutisches Gelenkschutzprogramm im Vergleich zu einer anderen Therapie, z.B. einer medikamentösen Behandlungsmethode, aus ökonomischer Sicht sinnvoller ist. »Intervention« bedeutet nicht nur Behandlungsmethode, sondern es kann sich auch z.B. um die ökonomische Analyse zweier Erfassungsinstrumente handeln.

Wie unterscheiden sich nun ökonomische Studien von anderen wissenschaftlichen Studien? Sind sie grundsätzlich anders oder gibt es Gemeinsamkeiten? Diese Fragen werden am Beispiel der Wirksamkeitsstudien geklärt.

In gesundheitsökonomischen Studien geht es wie bei Wirksamkeitsstudien um die Frage, ob eine Intervention zu einem besseren Ergebnis als ohne Behandlung oder im Vergleich zu anderen Interventionen führt. Daher sind die Studienarten dieselben wie bei den Wirksamkeitsstudien, d.h. es handelt sich bei wissenschaftlich hochstehenden Studien um **randomisiert kontrollierte Studien** (► Buch; Mangold: Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie Kap. 18.2.1) und – als übergeordnete Studien – um **systematische Übersichtsartikel** (► Buch; Mangold: Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie: Kap. 21). Der Unterschied zwischen gesundheitsökonomischen Studien und z.B. Wirksamkeitsstudien liegt also nicht im Studiendesign, sondern in der **Art der Zielgrößen** und in deren **Auswertung**.

- **Gesundheitsökonomische Studien berücksichtigen zwangsläufig die Kosten, Wirksamkeitsstudien meistens nicht, dafür gehen letztere häufig detaillierter auf verschiedene gesundheitliche Aspekte ein. Eine Studie kann aber auch gleichzeitig eine Wirksamkeitsstudie und eine gesundheitsökonomische Studie sein.**

Auch bei gesundheitsökonomischen Studien muss der Therapeut abklären, ob sie wissenschaftlich wertvoll, also valide sind, ob die Resultate bedeutsame Unterschiede zwischen verschiedenen Interventionen oder zwischen einer Intervention und keiner Intervention ergeben und ob sich die Ergebnisse in die Praxis umsetzen lassen. Diese Fragen werden nachfolgend abgehandelt.

Praxistipp

Es gibt eine Datenbank zu ökonomischen Studien, die **National Health Service Economic Evaluation Database (NHSEED)**. Sie ist unter ► <http://www.crd.york.ac.uk/crdweb/Home.aspx?DB=NHS%20EED> aufzurufen und enthält bibliografische Angaben und Bewertungen zu ökonomischen Studien, welche Kosten und Zielgrößen von Therapie- und Pflegealternativen im Gesundheitswesen miteinander vergleichen.

8 Ist die Evidenz zu einer gesundheitsökonomischen Analyse valide?

Die Validität gesundheitsökonomischer Analysen wird anhand verschiedener Leitfragen beurteilt. Die Übersicht listet die Leitfragen auf. Im Text werden sie nochmals einzeln aufgeführt und erklärt.

Leitfragen zur Beurteilung der Validität gesundheitsökonomischer Studien

- Ergibt die Analyse einen vollen ökonomischen Vergleich zwischen alternativen Strategien der Gesundheitsversorgung? (Referenzen: a, b)
- Gibt es Angaben darüber, aus welcher Perspektive Kosten und Nutzen analysiert wurden? (Referenzen: a, b)

- Falls sich die ökonomische Studie auf andere wissenschaftliche Studien bezieht: Sind diese valide? (Referenzen: b)
 - Wurden wichtige Messgrößen zur Beschreibung und geeignete Verfahren zur Messung und Bewertung des Gesundheitszustandes gewählt? (Referenzen: a, b)
 - Wurden alle relevanten Kosten aufgeführt? (Referenzen: a, b)
 - War das Follow-up ausreichend lang? (Referenzen: a, b)
 - Wurde eine Sensitivitätsanalyse durchgeführt? (Referenzen: a, c)
 - Wurden Kosten und Nutzen diskontiert? (Referenzen: a)
 - Wurden Untergruppen mit verschiedenen Baseline-Risiken unterschieden und deren ökonomische Analyse getrennt aufgeführt? (Referenzen: a)
- (Referenzen: a) Drummond et al. 1997; b) Sackett et al. 1999, S. 79-80; c) Sackett et al. 1999, S. 111-113)

■ **Ergibt die Analyse einen vollen ökonomischen Vergleich zwischen alternativen Strategien der Gesundheitsversorgung?**

Wenn eine Publikation einfach nur die Kosten einer oder mehrerer Interventionen aufzeigt, so handelt es sich nicht im eigentlichen Sinne um eine ökonomische Studie, sondern um eine Kostenanalyse. Eine ökonomische Analyse erfordert, mindestens 2 verschiedene Interventionen, Programme oder Strategien (auch z.B. eine Intervention mit keiner Intervention) hinsichtlich der Kosten und dem gesundheitlichen Nutzen **zu vergleichen**. Idealerweise berücksichtigt sie dabei alle klinisch-therapeutischen Methoden zur Behandlung der betreffenden Gesundheitsbeeinträchtigung.

Beispielsweise würde es also für eine ökonomische Analyse nicht ausreichen, bei der Behandlung eines Tennisellbogens (► Buch: 6 Kap. 18.5) lediglich die Kosten für Kortisonspritzen (Behandlung) und die Kosten für die Physiotherapie (Alternativbehandlung) aufzulisten. Vielmehr müsste auch der Nutzen in den Vergleich einbezogen werden sowie Kosten und Nutzen beider Behandlungsstrategien direkt verglichen bzw. miteinander verrechnet werden (► Abschn. 9).

■ **Gibt es Angaben darüber, aus welcher Perspektive Kosten und Nutzen analysiert wurden?**

Die Angaben darüber, aus welcher Perspektive die Studie durchgeführt wurde, sind wichtig, denn dann wird transparent, welche **Schwerpunkte** die Studie hinsichtlich der Kosten und Nutzen gesetzt hat. Folgende wesentliche Perspektiven lassen sich unterscheiden (Bühlren 2001):

- die **gesamtgeseellschaftliche Perspektive**, welche den gesamten Ressourcenverbrauch und alle Effekte berücksichtigt,
- die Perspektive des **Kosten-/Leistungsträgers**, die sich auf relevante Kostenfaktoren auf Seiten des Leistungsträgers beschränkt,
- die Perspektive der **Leistungserbringer**, in der v.a. betriebliche Kosten und Einnahmen relevant sind,
- weitere Perspektiven wie beispielsweise **Arbeitgeber, Patienten oder Angehörige**.

6 Bei Verweisen auf »Buch« ist immer Mangold (2013): Evidenzbasiertes Arbeiten in der Physio-/Ergotherapie, 2. Auflage gemeint

Die gesellschaftliche Perspektive ist die umfassendste Perspektive. Dort sind alle Kosten relevant, egal, wer sie zu tragen hat.

■ **Falls sich die ökonomische Studie auf andere wissenschaftliche Studien bezieht: Sind diese valide?**

Gesundheitsökonomische Studien können Daten anderer Studien, beispielsweise diejenigen randomisiert kontrollierter Studien, aufgreifen und in die eigenen Analysen mit einbeziehen. Diese Studien, auf welche die ökonomische Studie aufbaut, müssen ihrerseits valide sein. In der ökonomischen Studie sollte erkennbar sein, wie valide diese zugrunde liegenden Studien sind. Anhaltspunkte dafür, ob geeignete Beurteilungskriterien vorliegen, bieten die Leitfragen der Buchkapitel, z.B. für Wirksamkeitsstudien (► Buch: Kap. 18), für Erfassungsinstrumente (► Buch: Kap. 12, ► Buch: Kap. 13, ► Buch: Kap. 14, ► Buch: Kap. 15, ► Buch: Kap. 16, ► Buch: Kap. 17) und für systematische Übersichtsartikel bzw. Meta-Analysen (► Buch: Kap. 21).

■ **Wurden wichtige Messgrößen zur Beschreibung und geeignete Verfahren zur Messung und Bewertung des Gesundheitszustandes gewählt?**

Anhand einer oder verschiedener Messgrößen erfassen spezifische oder generische Erfassungsinstrumente den Gesundheitszustand und wandeln diesen – bei der Kosten-Nutzwert-Analyse – mittels spezifischer Bewertungsverfahren in Nutzwerte um. Diese Messgrößen müssen **für die Patienten wichtig** sein, um den Gesundheitszustand widerzuspiegeln, und die Umwandlungsverfahren in Nutzwerte müssen geeignet sein.

Hinweise geben die Erläuterungen zu den Erfassungsinstrumenten und Bewertungsverfahren weiter oben in diesem Text, um diese Leitfrage zu beurteilen. Ein weiterer Anhaltspunkt ist, ob die Wissenschaftler validierte Erfassungsinstrumente benutzten und fundierte Umwandlungsverfahren verwendeten. Ob die Messgrößen sinnvoll waren und das Spektrum der wichtigen Faktoren, die den Gesundheitszustand ausmachen, hinreichend abdeckten, lässt sich ergänzend dazu mit dem eigenen klinisch-therapeutischen Expertentum beurteilen.

■ **Wurden alle relevanten Kosten aufgeführt?**

Alle Kosten, welche für die festgelegte Perspektive (► 2. Leitfrage) relevant sind, müssen aufgeführt sein, sonst ist die Analyse unvollständig. Gerade aus gesellschaftlicher Perspektive bedeutet diese Forderung eine sehr **umfassende Kostenaufstellung**. Es ist wichtig, dass nicht nur monetäre Aufwendungen wie z.B. Kosten für das medizinisch-therapeutische Personal für die Durchführung der Intervention in der Studie berücksichtigt sind, sondern auch weitere Aufwendungen wie z.B. Kosten aufgrund des Arbeitsausfalls des Patienten oder allfälliger Behandlungen von Nebenwirkungen.

Für die spätere Frage nach der Anwendbarkeit ist wichtig, dass transparent wird, **unter welchen Bedingungen** die Kosten entstanden sind. Hierzu zählen sowohl das klinisch-therapeutische Setting als auch die lokalen Bedingungen wie z.B. gängige Tarife für das Personal, Medikamente, Hilfsmittel etc., denn die Kosten variieren von Land zu Land. Bei genügender Transparenz können der Leser und die Leserin einer ökonomischen Studie den Kostenfaktor für die eigenen Gegebenheiten korrigieren.

■ **War das Follow-up ausreichend lang?**

Es ist wichtig, dass das Follow-up genügend lang ist, damit die Wissenschaftler eine langfristige gesundheitsökonomische Bilanz ziehen können.

Welche Zeitspanne sinnvoll ist, hängt von der Fragestellung ab. Ist beispielsweise eine wiedergewonnene Arbeitsfähigkeit und damit Produktivität ein wichtiger Faktor der ökonomischen

Analyse, so darf die Zeitspanne nicht zu kurz sein, weil sonst das Einkommen auf einen kleinen Zeitraum beschränkt bleibt. Bei tödlich verlaufenden Krankheiten wählen die Wissenschaftler häufig den Tod als Endzeitpunkt. Näheres dazu, ► Buch: Kap. 7.4.1.

■ Wurde eine Sensitivitätsanalyse durchgeführt?

Ökonomische Daten sind häufig unvollständig und aus folgenden Gründen mit Unsicherheiten behaftet:

- Die Berechnungen basieren auf **Grundannahmen**. Beispielsweise schätzt die Studie, wie häufig Komplikationen, welche Kosten verursachen, durch die Behandlung auftreten. Ist die Annahme falsch, so stimmen die angegebenen Kosten nicht.
- Die Studie **vereinfacht** die komplexe Realität.
- Allgemein können **Messungenauigkeiten** auftreten.
- Die ökonomische Analyse kommt unter bestimmten **augenblicklichen Umständen** zustande. Diese äußeren Umstände können sich schnell ändern, z. B. durch Änderungen der Medikamentenkosten, Personalkosten oder der Beiträge, welche die Leistungsträger für eine Behandlung zu zahlen bereit sind.

Aufgrund dieser Unsicherheiten sollten die Autoren und Autorinnen einer Studie überprüft haben, wie **robust** ihre Analyse ist, d. h. wie konstant die Aussagen sind, wenn die Kosten innerhalb einer gewissen Bandbreite variieren. Diese Überprüfung heißt Sensitivitätsanalyse. Vereinfacht ausgedrückt dient die Sensitivitätsanalyse dazu herauszufinden: was wäre, wenn...?

Vorgehen bei der Sensitivitätsanalyse. Bei der Sensitivitätsanalyse variieren die Wissenschaftler wichtige Parameter, d. h. sie treffen Annahmen über deren **minimale und maximale Werte**. Beispielsweise nehmen sie an, dass die Personalkosten, die in der ursprünglichen Analyse auf 2.000 € veranschlagt wurden, um 10 % variieren und somit zwischen 1.800 und 2.200 € liegen könnten. Mit diesen Eckpunkten, d. h. den unteren und oberen Werten der angenommenen Bandbreite, führen die Autoren und Autorinnen einer Studie eine erneute ökonomische Berechnung durch und beurteilen mithilfe eines mathematischen Verfahrens, ob der Parameter sensitiv ist, d. h. ob dessen Variation etwas an der ursprünglichen Aussage ändert.

Das vorhergehende Beispiel bezog sich auf die Kosten der Intervention. Die Sensitivitätsanalyse berücksichtigt aber nicht nur Parameter der Kostenseite, sondern auch diejenigen der Nutzenseite.

Die Sensitivitätsanalyse ist ein wichtiges Instrument, um die Robustheit der ökonomischen Ergebnisse zu überprüfen. Jedoch darf nicht übersehen werden, dass die Auswahl der Größen, die variiert werden, und die angenommene Bandbreite der Schwankungen den **subjektiven Einschätzungen** der Untersucher unterliegen (Walshe u. Diehl 1998). Wenn der Nutzen auf experimentell gewonnenen Daten basiert, muss der Untersucher allerdings nicht selbst Annahmen über eine plausible Bandbreite treffen, sondern kann Konfidenzintervalle (► Buch: Kap. 7.5.1) verwenden (O'Brien et al. 1997).

■ Wurden Kosten und Nutzen diskontiert?

Studien ziehen sich häufig zumindest über mehrere Monate, wenn nicht sogar über Jahre hinweg. Entstehen die Kosten zweier Interventionen, die man einem ökonomischen Vergleich unterzieht, nicht ungefähr zum selben Zeitpunkt, so bewirkt das eine Verzerrung der Daten.

Beispiel Schädel-Hirn-Trauma

Für Patienten mit einem Schädel-Hirn-Trauma bietet ein ambulantes Rehabilitationszentrum 2 Arten kognitiven Trainings zur Auswahl an: ein computergestütztes Training, welches der Patient als Heimprogramm selbstständig durchführen kann (Intervention A) und ein therapeutengestütztes Training, bei welchem der Patient 1 Jahr lang regelmäßig zur Therapie ins Rehabilitationszentrum kommt (Intervention B).

Die Kosten der Intervention A fallen gleich zu Anfang an, da der Therapeut zu Beginn das gesamte Material (Computer, -software, -zubehör) für den Patienten anschaffen und ihn schulen muss. Danach führt der Patient die Therapie zu Hause selbstständig fort und sucht das Zentrum später nur noch zur Kontrolle oder bei Schwierigkeiten auf. Die Kosten der Intervention B verteilen sich dagegen gleichmäßig über 1 Jahr.

Der Kostenträger der Intervention B könnte dasjenige Geld, das erst später für die Therapie anfällt, gewinnbringend bei einer Bank anlegen. Müsste für beide Therapien der genau gleiche Betrag gezahlt werden, so wäre die Intervention B daher ökonomisch günstiger.

Um dieser Verzerrung entgegenzuwirken, sollten die Wissenschaftler die Kosten bei ökonomischen Studien **diskontieren**, d. h. mit einer Diskontierungsrate versehen, inhaltlich vergleichbar mit einem **Zinssatz**. Beispielsweise ist dann in der Studie zu lesen: »Die Kosten wurden mit einer jährlichen Rate von 3 % diskontiert«.

Wie hoch sollten die **Diskontierungsraten** aber sein? Smith u. Cravelle (2001) führten eine Literaturstudie über Empfehlungen zur Diskontierung im Gesundheitsbereich durch und fanden 16 Literaturquellen. Alle Quellen empfahlen eine Diskontierung sowohl der Kosten als auch der Gesundheitseffekte (Zielgrößen). Die Bandbreite der spezifischen Diskontierungsraten erstreckte sich von 1–8 % pro Jahr. Die am häufigsten genannten Raten lagen zwischen **3–5 % pro Jahr**. Die meisten Quellen empfahlen eine Sensitivitätsanalyse, indem die Diskontierungsraten variiert werden. Keine der Quellen empfahl, dass die Rate von der Länge des Zeithorizontes abhängen sollte. Nur eine Quelle empfahl, die Rate für die Gesundheitseffekte tiefer als die der Kosten anzusetzen.

Eine oft zitierte Quelle ist die CADTH (Canadian Agency for Drugs and Technologies in Health 2006). Sie empfiehlt eine Diskontierungsrate bei den zukünftigen Zielgrößen und Kosten von 5 % pro Jahr. Zudem soll diese Rate in einer Sensitivitätsanalyse variiert werden, beginnend bei 0 % pro Jahr (d. h. keine Diskontierung).

■ Wurden Untergruppen mit verschiedenen Baseline-Risiken unterschieden und deren ökonomische Analyse getrennt aufgeführt?

Ein unterschiedliches Baseline-Risiko verschiedener Untergruppen eines Krankheitsbildes beeinflussen die ökonomischen Ergebnisse z. T. beträchtlich (O'Brien et al. 1997). Untergruppen können sich beispielsweise durch spezifische Altersstufen, Schweregrade des Krankheitsbildes und Lokalisation der anatomischen oder physiologischen Störung unterscheiden. Aufgrund der Beeinflussung dieser Faktoren auf die ökonomische Analyse müssen diese Untergruppen getrennt aufgeführt sein.

Ist der Therapeut zum Urteil gelangt, dass die gesundheitsökonomische Studie valide ist, so muss er als nächsten Schritt herausfinden, ob die Ergebnisse für die Praxis bedeutsam sind. Wie er dabei vorgehen soll, erklären die folgenden Abschnitte.

■ Tab. 3 Kosten-Effektivitäts-Quotient (CER) und inkrementelles Kosten-Effektivitäts-Verhältnis (ICER) am Beispiel der Schmerzreduktion

	Intervention ₁	Intervention ₂	Differenz (= Increment)
Mittlere Kosten pro Patient (€)	4.000	2.000	2.000
Mittlere Wirksamkeit pro Patient (Anzahl Punkte)	3	1	2
CER (€/Punkt)	=4.000/3=1.333	=2.000/1=2.000	—
ICER Intervention ₁ vs. Intervention ₂ (€/Punkt)	—	—	=2.000/2=1.000

Die Wirksamkeit dieses fiktiven Beispiels bezieht sich auf Schmerzen, welche mit einer Schmerzskala von 0 (kein Schmerz) bis 10 (maximal vorstellbarer Schmerz) erfasst werden. Als Wirksamkeit ist in der Tabelle die mittlere Schmerzreduktion pro Patient aufgeführt. Je höher der Wert ist, desto größer ist die Wirksamkeit.

Die ICER, also der zusätzliche finanzielle Aufwand pro zusätzlichem Punkt Schmerzreduktion, der durch den Wechsel von der Intervention₂ zur Intervention₁ erreicht würde, beträgt 1.000 €.

9 Sind die Ergebnisse der gesundheitsökonomischen Analyse bedeutsam?

Ökonomische Studien dienen dazu, die Kosten und den gesundheitlichen Nutzen verschiedener Interventionen bzw. einer Intervention mit keiner Intervention zu vergleichen. Bei der Frage nach der Bedeutsamkeit der Ergebnisse gilt es herauszufinden, ob sich die ökonomischen Ergebnisse genügend unterscheiden, um eine Intervention aus alternativen Vorschlägen zu bevorzugen.

Zur Beantwortung der Frage stellt die Studie Kosten und Nutzen jeder Intervention dar und führt damit spezifische Berechnungen, die weiter unten beschrieben sind, durch. Außerdem zeigen gute Studien **Bandbreiten** auf, in welchen sich Kosten und Nutzen bewegen. Aus dem Vergleich dieser Berechnungen lässt sich schließen, ob das Ergebnis aus ökonomischer Sicht praxisrelevant ist.

Die Berechnungen sind in gesundheitsökonomischen Studien bereits durchgeführt und dargestellt. Zum besseren Verständnis werden sie trotzdem als Nächstes erklärt, gefolgt von den Leitfragen zur Untersuchung der Bedeutsamkeit der Evidenz.

Wie weiter oben beschrieben ist, lassen sich 3 wichtige Formen der gesundheitsökonomischen Studien unterscheiden, die Kosten-Effektivitäts-Analyse, Kosten-Nutzwert-Analyse und die Kosten-Nutzen-Analyse. Jede Form erfordert ihre eigenen, spezifischen Berechnungen, welche die folgenden Abschnitte separat aufführen.

9.1 Berechnungen der Kosten-Effektivitäts-Analyse (Cost-Effectiveness-Analysis)

Für den ökonomischen Vergleich mehrerer Interventionen gibt es bei der Kosten-Effektivitäts-Analyse, auch Kosten-Wirksamkeits-Analyse genannt, 2 Varianten: Die Berechnung des **Kosten-Effektivitäts-Quotienten** und des **inkrementellen (= zusätzlichen) Kosten-Effektivitäts-Verhältnisses**.

Kosten-Effektivitäts-Quotient

Der Kosten-Effektivitäts-Quotient, auch Kosten-Wirksamkeits-Quotient genannt, gibt an, wie viel eine Intervention pro Einheit klinisch-therapeutischer Nutzen im Schnitt pro Patient kostet. Der klinisch-therapeutische Nutzen ist der Unterschied zwischen den gesundheitlichen Veränderungen, die sich mit der Intervention ergeben, und denjenigen, die sich **ohne Intervention** ereignen. Der Nutzen ist hier also die **Wirksamkeit der Intervention**, wobei es sich wie bei den Kosten um einen Mittelwert der untersuchten Patientengruppe handelt. Die Art der Wirksamkeit kann z. B. aus Veränderungen der Punktzahl auf einer Schmerzskala oder ADL-Skala, der Anzahl gewonnener Lebens-

jahre, Anzahl erfolgreich behandelter Patienten oder auch physiologischen Größen wie Blutdruckveränderungen bestehen.

Um den **Kosten-Effektivitäts-Quotienten** (engl.: Cost-Effectiveness Ratio, **CER**) jeder Intervention auszurechnen, bilden die Wissenschaftler jeweils das Verhältnis aus den Kosten der Intervention zu ihrer Wirksamkeit (Bührlen 2001):⁷

$$\text{CER}_{\text{Intervention1}} = \frac{\text{Kosten}_1}{\text{Wirksamkeit}_1}$$

$$\text{CER}_{\text{Intervention2}} = \frac{\text{Kosten}_2}{\text{Wirksamkeit}_2}$$

$$\text{CER}_{\text{Intervention3}} = \frac{\text{Kosten}_3}{\text{Wirksamkeit}_3}$$

etc.

Die Intervention₁ ist dabei die zu beurteilende Intervention. Intervention₂, Intervention₃ etc. sind Vergleichsalternativen, also andere Behandlungen wie konventionelle Behandlungen und/oder andere neue Interventionen. Zur Verdeutlichung der Berechnung in der Zeile »CER«, ■ Tab. 3.

- – **Je kleiner der Kosten-Effektivitäts-Quotient ist, desto vorteilhafter ist die Intervention aus ökonomischer Sicht.**
- **Beim Vergleich der Ergebnisse verschiedener Interventionen gilt: Je mehr sich die Kosten-Effektivitäts-Quotienten unterscheiden, desto mehr lohnt sich der Umstieg zu derjenigen Intervention, welche den kleinsten Quotienten aufweist.**

Trotz dieser Vergleichsmöglichkeit sagt der Kosten-Effektivitäts-Quotient wenig aus, denn jede Intervention wird **für sich** berechnet. Wenn beispielsweise die Intervention₁ 5.000 € kostet und 25 Punkte Verbesserung im Selbsthilfestatus erbringt und die Intervention₂ 2.000 € kostet und die Leistung um 10 Punkte erhöht, so beträgt der Kosten-Effektivitäts-Quotient bei beiden Interventionen 200 €/Punkt. Für die Praxis ist jedoch wichtig, dass die 1. Intervention eine **größere Verbesserung** beim Patienten erreicht, wenn sie auch mehr kostet. Deshalb stellt sich die Frage, um wie viel sich der Patientennutzen erhöht, wenn man die Intervention₁ statt der Intervention₂ wählt, und wie viel das zusätzlich kostet. Genau das beinhaltet die nachfolgend beschriebene ökonomische Größe.

⁷ Achtung: CER wird in der Literatur unglücklicherweise oft als Synonym für ICER (Incremental Cost-Effectiveness Ratio) verwendet. Zudem ist die Abkürzung CER für Cost-Effectiveness Ratio nicht mit der CER (Control Event Rate) bei den anderen Themenbereichen (z. B. Wirksamkeitsstudien) zu verwechseln!

■ Tab. 4 Kosten-Nutzwert-Quotient (CUR) und inkrementelles Kosten-Nutzwert-Verhältnis (ICUR)

	Intervention ₁	Intervention ₂	Differenz (= Increment)
Mittlere Kosten pro Patient (€)	8.000	6.000	2.000
Mittlere Anzahl QALYs pro Patient (Anzahl QALYs)	0,8	0,5	0,3
CUR (€/QALY)	=8.000/0,8=10.000	=6.000/0,5=12.000	—
ICUR Intervention ₁ vs. Intervention ₂ (€/QALY)	—	—	=2.000/0,3=6.667

Inkrementelles Kosten-Effektivitäts-Verhältnis

Das inkrementelle Kosten-Effektivitäts-Verhältnis drückt aus, wie viel der Wechsel von einer Intervention zur anderen zusätzlich kostet und um wie viel der Patientennutzen dadurch steigt.

Inkrementell heißt zusätzlich. Die Zusatzkosten für eine Intervention im Vergleich zu der anderen heißen **inkrementelle Kosten**. Sie werden beim inkrementellen Kosten-Effektivitäts-Verhältnis ins Verhältnis zur zusätzlichen Nutzeneinheit gesetzt. Mit der **zusätzlichen Nutzeneinheit** ist z. B. die Anzahl Lebensjahre gemeint, welche ein Patient durch den Wechsel von der einen zur anderen Intervention gewinnen würde.

Ein Synonym für den Begriff inkrementelles Kosten-Effektivitäts-Verhältnis ist die **inkrementelle Kosten-Effektivitäts-Relation** und wird in den Studien und in der Literatur üblicherweise als **ICER** (engl.: Incremental Cost-Effectiveness Ratio) bezeichnet. Die Formel für die Berechnung lautet:

$$ICER = \frac{\text{Kosten}_1 - \text{Kosten}_2}{\text{Wirksamkeit}_1 - \text{Wirksamkeit}_2}$$

Kosten₁ sind dabei die Kosten der zu beurteilenden Intervention und Wirksamkeit₁ deren Wirksamkeit (z. B. Anzahl der durch die Intervention gewonnenen Lebensjahre) im Vergleich zu **keiner Intervention**. Kosten₂ und Wirksamkeit₂ sind entsprechend die Kosten und Wirksamkeit der Vergleichsalternative, also einer anderen Behandlung. Zur Verdeutlichung der Berechnung in der Zeile »ICER«, ■ Tab. 3.

- Eine hohe ICER bedeutet, dass hohe Kosten erforderlich sind, um durch den Wechsel von der Intervention₂ zur Intervention₁ eine Verbesserung um eine definierte Einheit (z. B. einen zusätzlichen Punkt auf einer Selbständigkeitsskala) zu erreichen.

9.2 Berechnungen der Kosten-Nutzwert-Analyse (Cost-Utility-Analysis)

Die Kosten-Nutzwert-Analyse ist ganz ähnlich wie die Kosten-Effektivitäts-Analyse. Der Unterschied liegt in der Einheit des Nutzens, welche hier das **QALY** (qualitätsangepasstes Lebensjahr) ist, ► Abschn. »Quality-adjusted-life-years (QALYs)«, ⁸

Auch bei der Kosten-Nutzwert-Analyse gibt es 2 wichtige Größen: den Kosten-Nutzwert-Quotienten und das inkrementelle Kosten-Nutzwert-Verhältnis (inkrementelle Kosten pro QALY).

Kosten-Nutzwert-Quotient

Der Kosten-Nutzwert-Quotient (engl.: Cost-Utility Ratio, CUR) gibt an, wie viel Geld eine Intervention im Schnitt pro Patient für ein QALY kostet.

Der Kosten-Nutzwert-Quotient berechnet sich aus dem Verhältnis der Kosten zu der Anzahl QALYs, welche durch die Intervention im Vergleich zu **keiner Intervention** gewonnen wird (Malek 2000, Bührlen 2001):

$$CUR_{\text{Intervention}_1} = \frac{\text{Kosten}_1}{\text{QALYs}_1}$$

$$CUR_{\text{Intervention}_2} = \frac{\text{Kosten}_2}{\text{QALYs}_2}$$

$$CUR_{\text{Intervention}_3} = \frac{\text{Kosten}_3}{\text{QALYs}_3}$$

etc.

Die Intervention₁ ist dabei die zu beurteilende Intervention. Intervention₂, Intervention₃ etc. sind Vergleichsalternativen, also eine herkömmliche Behandlung und/oder weitere neue Behandlungen. Zur Verdeutlichung der Berechnung, ■ Tab. 4.

- – Je kleiner der Kosten-Nutzwert-Quotient ist, desto vorteilhafter ist eine Intervention aus ökonomischer Sicht.
- Je mehr sich die Kosten-Nutzwert-Quotienten verschiedener Interventionen unterscheiden, desto mehr lohnt sich der Umstieg zu derjenigen Intervention, welche den kleinsten Quotienten aufweist.

Trotz dieser Vergleichsmöglichkeit sagt der Kosten-Nutzwert-Quotient wenig aus, denn jede Intervention wird für sich berechnet. Wenn beispielsweise die Intervention₁ 15.000 € kostet und 3 QALYs (im Vergleich zu keiner Intervention) erreicht, die Intervention₂ dagegen 5.000 € kostet und 1 QALY (im Vergleich zu keiner Intervention) erzielt, so liegt der Kosten-Nutzwert-Quotient bei beiden Interventionen bei 5.000 €/QALY. Aus Sicht des Kosten-Nutzwert-Quotienten unterscheiden sich also die beiden Interventionen nicht. Praxisrelevant ist jedoch, dass die 1. Intervention wesentlich **besser wirkt**, wenn sie auch mehr kostet. Deshalb stellt sich die Frage, um wie viel sich der Patientennutzen erhöht, wenn man die Intervention₁ statt der Intervention₂ wählt, und wie viel mehr man dafür bezahlen muss. Genau das beinhaltet die nachfolgend beschriebene ökonomische Größe.

Inkrementelles Kosten-Nutzwert-Verhältnis (inkrementelle Kosten pro QALY)

Das inkrementelle Kosten-Nutzwert-Verhältnis stellt die inkrementellen Kosten pro zusätzlicher Nutzeneinheit zweier Interventionen dar. **Inkrementelle Kosten** bedeuten die Zusatzkosten für eine Intervention im Vergleich zu der anderen. Mit **zusätzlicher Nutzeneinheit** ist die Anzahl QALYs gemeint, welche durch den Wechsel von der einen zur anderen Intervention gewonnen würde.

⁸ Grundsätzlich sind auch andere Größen als das QALY möglich, welche den Nutzen einer Intervention qualitativ und quantitativ möglichst umfassend beschreiben und sich an der Lebensqualität orientieren, aber das QALY hat sich international bei der Kosten-Nutzwert-Analyse durchgesetzt.

■ Tab. 5 Kosten-Nutzen-Quotient (CBR) und inkrementelles Kosten-Nutzen-Verhältnis (ICBR)

	Intervention ₁	Intervention ₂	Differenz (=Increment)
Mittlere Kosten pro Patient (€)	8.000	6.000	2.000
Mittlerer Nutzen pro Patient (€)	16.000	9.000	7.000
CBR (€/€)	=8.000/16.000=0,5	=6.000/9.000=0,67	—
ICBR Intervention ₁ vs. Intervention ₂ (€/€)	—	—	=2.000/7.000=0,29



■ Abb. 2 Ökonomie in der Therapie

Ein Synonym für den Begriff inkrementelles Kosten-Nutzen-Verhältnis ist die inkrementelle Kosten-Nutzen-Relation, üblicherweise **ICUR** (engl.: Incremental Cost-Utility Ratio) genannt. Die Formel für die Berechnung lautet:

$$ICUR = \frac{\text{Kosten}_1 - \text{Kosten}_2}{\text{QALY}_1 - \text{QALY}_2}$$

Kosten₁ sind dabei die Kosten der zu beurteilenden Intervention und QALY₁ deren Anzahl gewonnener qualitätsangepasster Lebensjahre im Vergleich zu **keiner Intervention**. Kosten₂ und QALY₂ sind entsprechend die Größen der Vergleichsalternative.

Zur Verdeutlichung der Berechnungen, ■ Tab. 4. Bei diesem Beispiel beträgt die ICUR, d.h. der zusätzliche finanzielle Aufwand für ein zusätzliches Lebensjahr mit perfekter Gesundheit durch den Wechsel von Intervention₂ zur Intervention₁, 6667€.

➤ **Eine hohe ICUR bedeutet, dass hohe Kosten erforderlich sind, um ein zusätzliches QALY durch den Wechsel von der Intervention₂ zur Intervention₁ zu gewinnen.**

Die ICUR gibt Auskunft über die Differenz der Kosten zweier Interventionen zur Differenz der Anzahl zusätzlicher QALYs pro Patient. Ob dieser Betrag akzeptabel ist, d.h. wo die Schmerzgrenze liegt, wie viel ein zusätzliches QALY kosten darf, wird weiter unten in der Leitfrage »Welche Empfehlung lässt sich aus den ökonomischen Werten und Kenngrößen ableiten?« diskutiert.

Beispiel Kosteneffektivität 1

Die Studie von Shaw et al. (2010) untersuchte die Wirksamkeit und Kosteneffektivität eines Therapieprogramms der oberen Extremität mit und ohne Behandlung mit Botulinumtoxin Typ A bei Patienten mit einer Spastizität in der oberen Extremität, verursacht durch einen Schlaganfall vor mindestens einem Monat.

Als Zielgröße der Kosteneffektivität wurden die Quality-adjusted life-years (QALYs) mithilfe des European Quality of Life-5 Dimensions (EQ-5D) für beide Gruppen bestimmt.

Die Studie kam zum Ergebnis, dass das inkrementelle Kosten-Nutzen-Verhältnis 93.500 £/QALY (Britische Pfund/QALY)⁹ beträgt. Die Autoren schlossen aus diesem hohen Betrag, dass die Behandlung mit Botulinumtoxin Typ A, welche zusätzlich zum Therapieprogramm der oberen Extremität zur Minderung der Spastizität in der oberen Extremität angewendet wird, bei dieser Patientengruppe nicht als kosteneffektiv angesehen werden kann.

9.3 Berechnungen der Kosten-Nutzen-Analyse (Cost-Benefit Analysis)

Die Kosten-Nutzen-Analyse untersucht die Verhältnisse der Kosten und Nutzen einer bzw. mehrerer Interventionen. Dabei sind sowohl Kosten als auch der Nutzen in monetären Werten ausgedrückt (Bührlen 2001).

Wie bei den vorhergehend beschriebenen Analysen gibt es auch hier 2 wichtige Größen: der Kosten-Nutzen-Quotient und das inkrementelle Kosten-Nutzen-Verhältnis.

Kosten-Nutzen-Quotient

Der Kosten-Nutzen-Quotient (engl.: Cost-Benefit Ratio, CBR) gibt das Verhältnis derjenigen Kosten, welche durch eine Intervention entstehen, zu denjenigen Kosten, die man aufgrund der Intervention im Vergleich zu **keiner Intervention** einspart (z. B. Invaliditätsrente), an. Sie beziffern dabei den Durchschnittswert pro Patient. Der Kosten-Nutzen-Quotient berechnet sich folgendermaßen (Roberts et al. 2000):

$$CBR_{\text{Intervention1}} = \frac{\text{Kosten}_1}{\text{Nutzen}_1}$$

$$CBR_{\text{Intervention2}} = \frac{\text{Kosten}_2}{\text{Nutzen}_2}$$

$$CBR_{\text{Intervention3}} = \frac{\text{Kosten}_3}{\text{Nutzen}_3}$$

etc.

Die Intervention₁ ist dabei die zu beurteilende Intervention. Intervention₂, Intervention₃ etc. sind Vergleichsalternativen, also eine herkömmliche Behandlung und/oder weitere neue Behandlungen.

Die Einheit sowohl der Kosten als auch des Nutzens ist die **Währung**, z. B. Euro oder USD. Zur Verdeutlichung der Berechnungen, ■ Tab. 5.

➤ **Je kleiner das Verhältnis zwischen Kosten und Nutzen ist, desto vorteilhafter ist die Intervention aus ökonomischer Sicht.**

⁹ 10.000 £ (Britische Pfund) entsprechen je nach momentanem Kurswert ca. 14.000-15.000 Euro.

- **Beim Vergleich der Verhältnisse mehrerer Interventionen gilt: Je mehr sich die Verhältnisse unterscheiden, desto mehr lohnt sich der Umstieg zu derjenigen Intervention, welche das kleinste Verhältnis aufweist.**

Aussagekräftiger für den ökonomischen Vergleich verschiedener Interventionen sind jedoch die **inkrementellen** Kosten pro Nutzen, das sog. inkrementelle Kosten-Nutzen-Verhältnis. Es stellt dar, um wie viel sich der Nutzen in Geldeinheiten erhöht oder verkleinert, wenn man die Intervention₁ statt der Intervention₂ wählt, und wie viel mehr oder wie viel weniger man dafür bezahlen muss.

Inkrementelles Kosten-Nutzen-Verhältnis

Das inkrementelle Kosten-Nutzen-Verhältnis stellt die inkrementellen Kosten pro zusätzlicher Nutzeneinheit zweier Interventionen dar. **Inkrementelle Kosten** sind die Zusatzkosten für eine Intervention im Vergleich zu einer anderen. Mit **zusätzlicher Nutzeneinheit** ist der monetäre Betrag gemeint, welchen man durch den Wechsel von der einen zur anderen Intervention einspart, z. B. durch die Vermeidung einer späteren Operation.

Das inkrementelle Kosten-Nutzen-Verhältnis (engl.: **Incremental Cost-Benefit Ratio**) wird üblicherweise **ICBR** abgekürzt. Die Formel für die Berechnung lautet:

$$\text{ICBR} = \frac{\text{Kosten}_1 - \text{Kosten}_2}{\text{Nutzen}_1 - \text{Nutzen}_2}$$

Kosten₁ sind dabei die Kosten der zu beurteilenden Intervention und Nutzen₁ der Geldbetrag, den man durch die Anwendung der Intervention₁ im Vergleich zu **keiner Intervention** spart. Kosten₂ und Nutzen₂ sind entsprechend die Größen der Vergleichsalternative.

Zur Verdeutlichung der Berechnungen, ■ Tab. 5. Die ICBR, d.h. der zusätzliche finanzielle Aufwand für den zusätzlichen Nutzen (monetären Gewinn) durch den Wechsel von Intervention₂ zur Intervention₁, beträgt bei diesem Beispiel 0,29.

- **Wenn ICBR < 1, so schneidet Intervention₁ ökonomisch besser als Intervention₂ ab.**

Grenzwerte oder ähnliche Interpretationshilfen gibt es nach Wissen der Autorin nicht.

9.4 Leitfragen zur Beurteilung der Bedeutsamkeit der Ergebnisse

Die vorhergehenden Abschnitte beschrieben, welche Berechnungen sich in gesundheitsökonomischen Studien finden lassen. Jedoch muss man nicht bei allen Studien diese Kenngrößen wie beispielsweise die ICER genauer betrachten, um zu entscheiden, ob die Evidenz für die Praxis bedeutsam ist. Wenn beispielsweise die Intervention₁ weniger kostet und gleichzeitig einen größeren Nutzen als die Intervention₂ aufweist, lässt sich die Intervention₁ empfehlen, ohne den genauen Wert für die ICER zu analysieren. Wenn aber – und dies ist ein häufiger Fall – z. B. die Intervention₁ zwar einen größeren Nutzen als die Intervention₂ aufweist, aber auch mehr kostet, so fragt sich, wo die finanzielle Schmerzgrenze pro Nutzeneinheit liegt, um die Intervention₁ empfehlen zu können.

Die Übersicht listet Leitfragen auf, welche zur Beurteilung der Bedeutsamkeit der Ergebnisse dienen.

Leitfragen zur Beurteilung der Bedeutsamkeit der Ergebnisse gesundheitsökonomischer Studien

- Wie hoch sind Kosten und Nutzen bzw. deren inkrementelle Werte und Verhältnisse?
(Referenzen: a, b)
 - Welche Empfehlung lässt sich aus den ökonomischen Werten und Kenngrößen ableiten?
(Referenzen: a)
 - Unterscheiden sich die inkrementellen Kosten und der inkrementelle Nutzen bzw. deren Verhältnisse zwischen Untergruppen?
(Referenzen: a)
 - Wie stark ändern sich die Ergebnisse bzw. Schlussfolgerungen aufgrund der Sensitivitätsanalyse?
(Referenzen: a, b)
- (Referenzen: a) O'Brien et al. 1997; b) Sackett et al. 1999, S. 111-113)

- **Wie hoch sind Kosten und Nutzen bzw. deren inkrementelle Werte?**

Diese und die nächste Leitfrage bilden **Kernfragen** der ökonomischen Analyse.

Um diese Leitfrage zu beantworten, sieht sich der Therapeut die Kosten und den Nutzen sowie die **Kenngrößen** ICER, ICUR oder ICBR an. Die Darstellungen der bloßen Zahlen für Kosten und Nutzen und der Kenngrößen führen aber noch nicht zu einer Entscheidung. Dazu sind **Interpretationshilfen** notwendig, wann es sich aufgrund der Kenngrößen empfiehlt, von einer Intervention zu einer anderen zu wechseln. Diese Hilfen liefert die nächste Leitfrage.

- **Welche Empfehlung lässt sich aus den ökonomischen Werten und Kenngrößen ableiten?**

Die in der vorhergehenden Leitfrage erwähnten und weiter oben beschriebenen ökonomischen Kenngrößen muss der Therapeut interpretieren, um Empfehlungen ableiten zu können. Wie geht er dabei vor? Die Erklärungen zu dieser Frage berücksichtigen zur Vereinfachung hier nur 2 Interventionen, nämlich die Intervention A und Intervention B.

Kosten und Nutzen lassen sich beim Vergleich zweier Interventionen in höher, gleich hoch oder niedriger einteilen: Die Intervention A kann im Vergleich zur Intervention B höhere, gleich hohe oder niedrigere Kosten verursachen und einen höheren, gleich hohen oder niedrigeren Nutzen aufweisen. Daraus entstehen 9 mögliche **Kombinationen**, welche zu starker, schwacher oder keiner **Dominanz für Entscheidungen** zugunsten einer Intervention führen (O'Brien et al. 1997). Zu den verschiedenen Kombinationen, ■ Tab. 6.

Einfach ist die Entscheidung zugunsten einer Intervention dann, wenn z. B. die Intervention A weniger kostet und gleichzeitig einen höheren Nutzen als die Intervention B aufweist (■ Tab. 6, Kombination 1). Drei der 9 möglichen Kombinationen lassen jedoch offen, welche Intervention zu bevorzugen ist. Welche Überlegungen helfen in diesen Fällen bei der Entscheidungsfindung?

- **Kombination 7** (■ Tab. 6): Hier herrscht **Neutralität**, denn beide Interventionen verursachen gleiche Kosten und weisen denselben Nutzen auf. In diesem Fall muss der Therapeut überlegen bzw. nachforschen, ob andere Gründe vorliegen, um eine der Interventionen zu bevorzugen. Gründe könnten hier beispielsweise Nebenwirkungen, Vertrautheit mit der Intervention oder der Zugang zu den benötigten Mitteln (Material, Fortbildungen, Medikamente etc.) sein. Vielleicht sind bestimmte Gründe bereits in der Studie selbst angegeben.

■ Tab. 6 Entscheidungshilfen für verschiedene Kombinationen von Kosten und Nutzen zweier Interventionen A und B

Kombination	Kosten	Nutzen	Entscheidung zugunsten von	Bemerkungen
1	A<B	A>B	A	Starke Dominanz für Entscheidung
2	A>B	A<B	B	Starke Dominanz für Entscheidung
3	A=B	A>B	A	Schwache Dominanz für Entscheidung
4	A<B	A=B	A	Schwache Dominanz für Entscheidung
5	A=B	A<B	B	Schwache Dominanz für Entscheidung
6	A>B	A=B	B	Schwache Dominanz für Entscheidung
7	A=B	A=B	?	Keine Dominanz: Kosten und Nutzen sind neutral. Gibt es andere Gründe, eine der Interventionen zu bevorzugen?
8	A>B	A>B	?	Keine Dominanz. Weitere ökonomische Abklärungen sind notwendig
9	A<B	A<B	?	Keine Dominanz. Weitere ökonomische Abklärungen sind notwendig

■ **Kombinationen 8 und 9** (■ Tab. 6): Hier stellt sich die Frage, ob der höhere Nutzen die höheren Kosten wert ist, um die teurere Intervention zu bevorzugen. Dazu zieht man die ökonomischen Kenngrößen wie das **inkrementelle Kosten-Nutzwert-Verhältnis** zur Rate. Als grobe Entscheidungshilfe orientiert man sich an Schwellenwerten für das inkrementelle Verhältnis, um zu entscheiden, ob die Intervention als kosteneffektiv und als gesundheitsökonomisch akzeptabel gelten kann. Allerdings gibt es bisher in **keinem Land** einen offiziell festgelegten, **verbindlichen Schwellenwert** für die maximal zulässige Kosteneffektivität eines medizinischen Verfahrens (Greiner 2007). Zudem sollen Kenngrößen wie das inkrementelle Kosten-Nutzwert-Verhältnis nicht einzige Entscheidungsgrundlage sein, wie weiter unten weiter ausgeführt wird.

Entscheidungsfindung verschiedener Institute des Gesundheitswesens. Institute in verschiedenen Ländern beschäftigen sich mit Entscheidungsfindungen über medizinische Verfahren, in denen gesundheitsökonomische Evaluationen häufig Teil eines umfassenden **Health Technology Assessments** sind (Greiner 2007). Health Technology Assessment (HTA) bezeichnet einen Prozess zur systematischen Bewertung medizinischer Technologien, Prozeduren und Hilfsmittel, aber auch Organisationsstrukturen, in denen medizinische Leistungen erbracht werden. Untersucht werden dabei Kriterien wie Wirksamkeit, Sicherheit und Kosten, jeweils unter Berücksichtigung sozialer, rechtlicher und ethischer Aspekte. Das Ergebnis einer HTA-Studie wird in der Regel als HTA-Bericht veröffentlicht. Dieser soll primär als Entscheidungshilfe bei gesundheitspolitischen Fragestellungen dienen. Dazu gehört unter anderem die Übernahme von Innovationen in den Leistungskatalog der Gesetzlichen Krankenversicherung durch den Gemeinsamen Bundesausschuss (G-BA, Wikipedia 2008).

Besonders intensiv wurde die Entscheidungsfindung des **National Institute for Clinical Excellence (NICE)** von Großbritannien analysiert und diskutiert. Der gängige – aber nicht absolute – Schwellenwert wird hier häufig im Bereich zwischen **20.000–30.000 £/QALY** angegeben¹⁰ (Devlin u. Parkin 2004, Appleby et al. 2007). Gründe, warum das NICE einen absoluten Schwellenwert für die Beurteilung der Akzeptanz einer Technologie im Gesundheitswesen ablehnt, liegen u. a. darin, dass es keine empirische Basis dafür gibt, wo die Schwelle angesetzt werden soll und dass Umstände bestehen können, warum das NICE

eine Schwelle ignorieren will (Rawlins u. Culyer 2004). Rawlins u. Culyer (2004) differenzierten 2 verschiedene Spannen und hielten fest: Es ist unwahrscheinlich, dass das NICE eine Technologie mit einem inkrementellen Kosten-Nutzwert-Verhältnis von 5.000–15.000 £/QALY nur aufgrund dieses ökonomischen Faktors ablehnen würde. Dagegen bräuhete es spezielle Gründe, sie mit einem inkrementellen Kosten-Nutzwert-Verhältnis über 25.000–35.000 £/QALY zu akzeptieren.

Das NICE entscheidet daher **von Fall zu Fall** (Rawlins u. Culyer 2004). Es entscheidet über Akzeptanz oder Ablehnung von Technologien im Gesundheitswesen nicht nur anhand des inkrementellen Kosten-Nutzwert-Verhältnisses, sondern auch anhand weiterer Faktoren (Appleby et al. 2007, Rawlins u. Culyer 2004, Devlin u. Parkin 2004):

- **Unsicherheit der Abschätzung** hinsichtlich der Kosten und des Nutzwertes,
- **spezielle Merkmale der Bedingungen und Population**, welche die Technologie nutzt,
- **innovativer Charakter** der Technologie,
- ggf. die weiter gefassten gesellschaftlichen Kosten und den Nutzen,
- ggf. Bezugnahme auf vorhergehende Einschätzungen.

Devlin u. Parkin (2004) analysierten 51 Entscheidungen des NICE über Akzeptanz oder Ablehnung von Technologien. Neben der ICER fanden sie als 2 wichtigste Faktoren bei der Entscheidungsfindung des NICE:

1. das Ausmaß der Unsicherheit der Abschätzung,
2. die Belastung der Krankheit, d. h. die Anzahl betroffener Patienten, auf welche sich die Technologie bezieht.

In Deutschland sind das Deutsche Institut für Medizinische Dokumentation (**DIMDI**) sowie das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (**IQWiG**) für HTA federführend (Greiner 2007).

Für den deutschsprachigen Bereich sind der Autorin dieser Publikation keine Angaben über Schwellenwerte bekannt.

Zusätzlich zu der Unverbindlichkeit der Schwellenwerte gibt es noch eine weitere Einschränkung: Während es sinnvoll und praktikabel ist, für gewonnene QALYs bzw. gewonnene Lebensjahre ökonomische Schwellenwerte einzusetzen, wird es bei Zielgrößen der Kosten-Effektivitäts-Analysen oder Kosten-Nutzen-Analysen schwieriger. Wo soll die Schwelle z. B. pro gewonnener Einheit an Selbständigkeit oder Schmerzreduktion liegen? Nach Wissen der Autorin bestehen hier noch keine Interpretations- oder Entscheidungshilfen.

10 10.000 £ (Britische Pfund) entsprechen je nach momentanem Kurswert ca. 14.000–15.000 Euro.

■ Unterscheiden sich die inkrementellen Kosten und der inkrementelle Nutzen bzw. deren Verhältnisse zwischen Untergruppen?

Untergruppen, z. B. Patientengruppen bestimmter Altersstufen desselben Krankheitsbildes, können unterschiedliche Baseline-Risiken und daher verschiedene ökonomische Ergebnisse aufweisen (O'Brien et al. 1997). Daher können ökonomische Entscheidungen für verschiedene Untergruppen unterschiedlich ausfallen. Die Vorgehensweise zur Beurteilung jeder Untergruppe entspricht derjenigen der vorhergehenden Leitfrage.

■ Wie stark ändern sich die Ergebnisse bzw. Schlussfolgerungen aufgrund der Sensitivitätsanalyse?

Die Ermittlung der Kosten und des Nutzens sind in den Studien mit einem Unsicherheitsfaktor behaftet. Deshalb sollten die Studien eine Sensitivitätsanalyse enthalten, welche Kosten und Nutzen in einer **sinnvollen Bandbreite** (bzw. anhand von Konfidenzintervallen) variiert und so Annahmen über deren minimale und maximale Werte trifft. Bei der Betrachtung der Ergebnisse sollte der Therapeut ein Augenmerk darauf richten, ob sich diese minimalen und maximalen Werte so stark von den angenommenen mittleren Werten und Berechnungen unterscheiden, dass die Schlussfolgerungen **zu ändern sind**.

Beispiel Kosteneffektivität 2

1. Der Schwellenwert liegt bei 40.000 €/QALY, abgeleitet aus dem Wert von Großbritannien (s. o.).
2. Das inkrementelle Kosten-Nutzwert-Verhältnis, welches die Studie anhand der mittleren angenommenen bzw. experimentell ermittelten Werte für Kosten und Nutzwerten berechnete, liegt bei 30.000 €/QALY.
3. Die Bandbreite des inkrementellen Kosten-Nutzwert-Verhältnisses erstreckt sich von 25.000–50.000 €/QALY.

Interpretation: Die Berechnung anhand der mittleren Werte ließe die Schlussfolgerung zu, dass die teurere Intervention als kosteneffektiv und als gesundheitsökonomisch akzeptabel gälte, da der Schwellenwert von 40.000 €/QALY nicht erreicht ist. Die Berechnungen mit den Minimal- und Maximalwerten zeigen jedoch, dass der obere Wert des inkrementellen Kosten-Nutzwert-Verhältnisses diesen Schwellenwert übersteigt. Für den Maximalwert ändert sich die Schlussfolgerung, die Kosteneffektivität kann angezweifelt werden.

Nachdem der Therapeut die Güte einer ökonomischen Studie bewertet und deren Ergebnisse hinsichtlich der Bedeutsamkeit ausgewertet hat, muss er überprüfen, ob sich die Evidenz auf seinen Patienten übertragen lässt. Die in der Studie ermittelten Kosten und Nutzen können nämlich beträchtlich von den in der Praxis zu erwartenden Kosten und Nutzen abweichen, z. B. »... wenn Studienpatienten aufgrund enger Kriterien rekrutiert werden oder etwa der Untersuchungs- und Therapieaufwand und die damit verbundenen Kosten oder auch Nutzen wesentlich höher sind.« (Walshe u. Diehl 1998).

10 Ist die Evidenz auf die medizinisch-therapeutische Versorgung des Patienten anwendbar?

Die in der Studie ermittelten Kosten und der Nutzen beziehen sich auf eine mehr oder weniger repräsentative Stichprobe der untersuchten Patientengruppe. Eventuell treffen die Ergebnisse auf den eigenen Patienten bzw. Behandlungsrahmen zu, vielleicht gibt es aber auch Abweichungen davon. Dieser Abschnitt hilft abzuschätzen, ob Kosten und Nutzen in der **Routineversorgung** in der gleichen Größenordnung wie in der Studie zu

erwartet sind. Diese Abschätzung erfolgt mithilfe der ersten 2 Leitfragen, welche die Übersicht aufführt und die weiter unten beschrieben sind.

Des Weiteren muss der Therapeut überlegen, ob die Anwendung der ökonomisch günstigeren Intervention in der eigenen Arbeitssituation überhaupt möglich ist. Dazu gibt es eine Leitfrage zu den Präferenzen des Patienten und eine zur praktischen Realisierbarkeit.

Leitfragen zur Beurteilung der Anwendbarkeit der Ergebnisse gesundheitsökonomischer Studien

- Kann der eigene Patient einen gleichartigen gesundheitlichen Nutzen erwarten?
(Referenzen: a)
- Kann der eigene Patient gleichartige Kosten erwarten?
(Referenzen: a)
- Werden das Therapieschema und seine Folgen den Wertvorstellungen und Präferenzen des Patienten gerecht?
(Referenzen: b)
- Wurde die Behandlung so genau beschrieben, dass sie nachgeahmt werden kann, und ist sie praxisnah?
(Referenzen: keine)

(Referenzen: a) O'Brien et al. 1997; b) Sackett et al. 1999, S. 132–133)

■ Kann der eigene Patient einen gleichartigen gesundheitlichen Nutzen erwarten?

Die Übertragbarkeit der Ergebnisse hinsichtlich des Nutzens hängt im Wesentlichen von 2 Faktoren ab, die zu überprüfen sind (O'Brien et al. 1997):

- Erstens muss man danach fragen, ob der eigene Patient den Studienpatienten **ausreichend ähnelt**. Wenn der eigene Patient den **Einschluss- und Ausschlusskriterien** der Studie entspricht, so ist die Ähnlichkeit gewährleistet. Um noch sicherer zu gehen, sollte man zusätzlich einen Blick auf die **Baseline-Eigenschaften** der tatsächlich in die Studie aufgenommenen Patienten werfen. Weicht der eigene Patient von der Studienpopulation ab, so ist zu überlegen, ob es wichtige Faktoren sind oder ob der gesundheitliche Nutzen trotz der Unterschiede in der gleichen Größenordnung liegen könnte. Weicht der Patient z. B. im Alter oder Schweregrad der Krankheit von der Studienpopulation zu stark ab oder weist er zusätzliche Krankheiten oder Komplikationen auf, so könnte er einen im Vergleich zur Studiengruppe erhöhten oder niedrigeren gesundheitlichen Nutzen erlangen. Es empfiehlt sich darauf zu achten, ob die Studie Untergruppen differenziert hat. Diejenige **Untergruppe**, welche dem eigenen Patienten am meisten ähnelt, sollte man dann zur Rate ziehen.
- Zweitens ist wichtig abzuklären, ob die **klinisch-therapeutische Versorgung** der Studienpatienten den eigenen Möglichkeiten und Bedingungen ausreichend ähnelt. Ist es beispielsweise möglich, den eigenen Patienten genauso früh nach der Diagnosestellung, über denselben Behandlungszeitraum – z. B. 3 Monate lang – und mit derselben Intensität wie die Patienten in der Studie zu behandeln? Wenn nicht, so kann sich das auf den gesundheitlichen Nutzen und auch auf die Kosten auswirken, wodurch die Übertragbarkeit der ökonomischen Ergebnisse der Studie auf den eigenen Patienten nicht mehr gewährleistet ist. Ein bedeutender Faktor für die Bedingungen der klinisch-therapeutischen Versorgung ist das **Land**, in welchem die Studie durchgeführt wurde und in welchem der Therapeut

selbst praktiziert. Das klinisch-therapeutische **Setting**, z. B. der ambulante oder stationäre Bereich, ist ein weiterer Faktor, den man bei der Analyse zu berücksichtigen hat.

Ramsey et al. (2001) nennen noch weitere Aspekte, welche sich sowohl auf den Nutzen als auch z. T. auf die Kosten auswirken und damit zu abweichenden Ergebnissen in der Praxis führen können:

- Die Vergleichsgruppe erhält häufig eine für die Praxis **untypische Behandlung**. Dadurch ist der ökonomische Vergleich der Intervention mit der Kontrollbehandlung praxisfern.
- Einschlussverfahren bewirken, dass Patienten in Studien **homogener** als in der Praxis sind (vgl. oben) und dass deren **Compliance** größer ist.
- Die Erfahrung der Fachleute in der Studie mit dem Krankheitsbild und mit den angewendeten Behandlungen spielt eine Rolle. Sind – wie häufig in den Studien – **erfahrene Fachleute** beteiligt, so fallen die Ergebnisse wahrscheinlich besser aus, als sie in der Praxis vorzufinden wären.
- Weil in der Studie **häufigere Kontrollen des Gesundheitszustandes** als in der Praxis stattfinden, erhöht sich die Wahrscheinlichkeit, gesundheitliche Probleme festzustellen.

Das Vorhandensein dieser vielen Einflussfaktoren soll nicht zu dem Schluss führen, dass es keine Übertragbarkeit der Studienergebnisse auf die Praxis geben kann. Vielmehr soll der Therapeut sich überlegen, welche Faktoren besonders wichtig sind. Diese sollten ausreichend übereinstimmen. Manche Faktoren lassen sich auch beeinflussen. Die Compliance beispielsweise kann der Therapeut in der Praxis vielleicht durch Aufzeigen der Studienergebnisse erhöhen. Auch die Erfahrungen bzw. Kenntnisse lassen sich durch Schulungen und mit der zunehmenden Anwendung der Intervention erhöhen.

■ Kann der eigene Patient gleichartige Kosten erwarten?

Zur Beantwortung dieser Frage ist zu berücksichtigen, dass die Kosten das Produkt aus der **Menge** der verbrauchten Ressourcen (z. B. Anzahl therapeutischer Behandlungen, Tests oder Medikamente) und deren **Preis pro Einheit** darstellen. Beide Faktoren können von der eigenen Praxis abweichen, denn die Möglichkeiten und Bestimmungen variieren u. a. von Land zu Land.

Bei der Durchführung von Studien gibt es im Vergleich zur Praxis **zusätzlichen Aufwand**. Zu beachten ist beispielsweise, ob Tests für das Einschlussverfahren, d. h. für die Überprüfung, ob der Patient für die Studie in Frage kommt, die Kosten künstlich erhöht haben (Ramsey et al. 2001).

Zudem ist zu berücksichtigen, dass die Interpretation der ökonomischen Ergebnisse v. a. vom **Land**, in dem der Therapeut praktiziert, abhängt: Beispielsweise gibt es keinen Grund, warum ein Kosten-Effektivitäts-Verhältnis von 50.000 USD pro gewonnenem Lebensjahr, welches einen akzeptablen Schwellenwert für die USA darstellt, auch für ein weniger industrialisiertes Land gelten sollte (O'Brien et al. 1997). Dieses Geld stünde anderen Projekten nicht mehr zur Verfügung, welche möglicherweise einen größeren Nutzen für die Bevölkerung hätten.

■ Werden das Therapieschema und seine Folgen den Wertvorstellungen und Präferenzen des Patienten gerecht?

Der Patient muss bereit sein, die Intervention aktiv zu unterstützen. Dazu ist es notwendig, dass die Art der Intervention mit deren **Nutzen und Risiken**, allfällige **Eigenkosten** der Therapie und der **zeitliche Aufwand** dem Patienten gerecht werden und

seinen persönlichen Zielen entsprechen. Der Therapeut oder der Arzt bereitet dazu die notwendigen Informationen anhand der wissenschaftlichen Literatur patientengerecht auf und vermittelt sie dem Patienten, sodass er sie mit seinen Wertvorstellungen und Präferenzen vergleichen und eine Entscheidung – häufig mithilfe weiterer Unterstützung durch die Fachperson – treffen kann.

■ Wurde die Behandlung so genau beschrieben, dass sie nachgeahmt werden kann, und ist sie praxisnah?

Für die Anwendung der Intervention sind Fachkompetenz und häufig bestimmte Geräte oder Materialien (therapeutisches Material, Hilfsmittel, Medikamente etc.) notwendig. Daher muss der Therapeut überlegen, ob er die Intervention schon anhand der eigenen Erfahrung und der Beschreibung in der Studie nachahmen kann. Beispielsweise ist eine Medikamentengabe durch einen Arzt sicher einfacher zu beschreiben und in die Praxis umzusetzen als eine komplexe Intervention. Wenn die Intervention neue Fachkompetenz erfordert, muss der Therapeut überlegen, ob er sie erwerben kann, z. B. in Form von Fortbildungen. Zu Überlegungen der Praktikabilität gehört auch abzuklären, ob er allfällige Geräte und Materialien beschaffen kann, ob sie finanzierbar sind und ob genügend Platz dafür zur Verfügung steht.

Literatur

- Appleby J, Devlin N, Parkin D (2007) NICE's cost effectiveness threshold. *BMJ* 335:358-359
- Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A (2007) Measuring and valuing health benefits for economic evaluation. Oxford University Press Inc., New York
- Brazier J, Roberts J, Deverill M (2002) The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 21(2):271-292
- Brazier J, Roberts J, Tsuchiya A, Busschbach J (2004) A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 13(9):873-884
- Bührlen B (2001) Konzepte und Methoden der ökonomischen Evaluation. Vorlesung »Grundlagen der Evaluation und Forschungsmethodik«, Hochtun-Institut für Rehabilitationsforschung Bad Säckingen, 2001. <http://www.buehrlen.de/downloads/oekev01f.pdf>. Zugriffen 23 April 2008
- Canadian Agency for Drugs and Technologies in Health (CADTH) (2006) Guidelines for the economic evaluation of health technologies. Canadian Agency for Drugs and Technologies in Health, Ottawa
- Devlin N, Parkin D (2004) Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13(5):437-452
- Dolan P (1997) Modeling valuations for EuroQol health states. *Med Care* 35:1095-1108
- Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D for the Evidence-Based Medicine Working Group (1997) Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA* 277(19):1552-1557
- Drummond MF, Sculpher M, O'Brien B, Stoddart GL, Torrance GW (2005) Methods for the economic evaluation of health care programmes. Oxford Medical Publications, Oxford
- Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M (2002) Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 40(2):113-128
- Greiner W (2007) Die Rolle der Pharmaökonomie zur Ressourcenallokation im Gesundheitswesen – Überblick und Implikationen für Deutschland. *Qual manag* 12:51-56
- Hawthorne G, Richardson J, Day NA (2001) A comparison of the assessment of quality of life (AQoL) with four other generic utility instruments. *Ann Med* 33(5):358-370

- Hawthorne G, Richardson J, Osborne R, McNeil H (1997) The Australian Quality of Life (AQoL) instrument: Initial validation. Working Paper 66, Centre for Health Program Evaluation
- Kaplan RM, Anderson JP (1988) A general health policy model: update and applications. *Health Serv Res* 23(2):203-235
- Klier H, Pogantsch M (2007) EBM-Eck: Pharmakoökonomische Studien. *GKK med info* 17(6):1-4
- Luijsterburg PAJ, Lamers LM, Verhagen AP, Ostelo RWJG, van den Hoogen HJMM, Peul WC, Avezaat CJJ, Koes BW (2007) Cost-effectiveness of physical therapy and general practitioner care for sciatica. *Spine* 32(18):1942-1948
- Malek M (2000) Implementing QALYs. What is...? series 2(1). Hayward Group plc
- O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF for the Evidence-Based Medicine Working Group (1997) Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA* 277(22):1802-1806
- Phillips C, Thompson G (2001a) What is cost-effectiveness? What is...? series 1(3). Hayward Group plc
- Phillips C, Thompson G (2001b) What is a QALY? What is...? series 1(6). Hayward Group plc
- Ramsey SD, McIntosh M, Sullivan SD (2001) Design issues for conducting cost-effectiveness analyses alongside clinical trials. *Annu Rev Public Health* 22:129-141
- Rawlins MD, Culyer AJ (2004) National Institute for Clinical Excellence and its value judgments. *BMJ* 329(7459):224-227
- Roberts PR, Betts TR, Morgan JM (2000) Cost effectiveness of the implantable cardioverter defibrillator. *Eur Heart J* 21:712-719
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB (1999) Evidenzbasierte Medizin – EBM-Umsetzung und –vermittlung. Deutsche Ausgabe: Kunz R, Fritsche L. Zuckschwerdt, München
- Shaw L, Rodgers H, Price C, van Wijck F, Shackley P, Steen N, Barnes M, Ford G, Graham L; BoTULS investigators (2010) BoTULS: a multicentre randomised controlled trial to evaluate the clinical effectiveness and cost-effectiveness of treating upper limb spasticity due to stroke with botulinum toxin type A. *Health Technol Assess* 14(26):1-113, iii-iv
- Sintonen H (1994) The 15-D Measure of Health Related Quality of Life: Reliability, Validity and Sensitivity of its Health State Descriptive System. Working Paper 41, Centre for Health Program Evaluation. <http://www.buseco.monash.edu.au/centres/che/pubs/wp41.pdf>. Zugegriffen 17 Dezember 2007
- Sintonen H, Pekurinen M (1993) A fifteen dimensional measure of health-related quality of life (15D) and its applications. In: Walker SR, Rosser RM. (eds) *Quality of life assessment. Key issues in the 1990s*. Kluwer, Dordrecht
- Smith DH, Cravalle H (2001) The practice of discounting in economic evaluations of healthcare interventions. *Int J Technol Assess Health Care* 17:236-243.
- The EuroQol Group (1990) EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 16:199-208
- Torrance GW (1982) Multi-attribute utility theory as a method of measuring social preferences for health states in long-term care. In: Kane RL, Kane RA (eds) *Values and long-term care*. Lexington Books, Lexington, MA
- Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q (1996) Multiattribute utility function for a comprehensive health status classification system. *Health Utilities Index Mark 2*. *Med Care* 34(7):702-722
- Walshe R, Diehl V (1998) Ökonomische Evaluationen im Rahmen klinischer Studien. *Internist* 39(9):943-954
- Walton J, Barondess JA, Lock S (1994). *The Oxford Medical Companion*. Oxford University Press, Oxford
- Wikipedia (2008) Health Technology Assessment http://de.wikipedia.org/wiki/Health_Technology_Assessment. Zugegriffen 12 Juli 2008

Zum Weiterlesen

Single-subject research design

Zusammenfassung

Das Single-subject research design bildet eine wissenschaftlich anerkannte Art von Studien an einzelnen Patienten mit dem Ziel, die Wirksamkeit einer Behandlung im Vergleich zu keiner Behandlung, einem Placebo oder zu einer anderen Behandlung nachzuweisen. Dazu braucht es im einfachsten Fall, dem AB Design, eine Baseline und eine Interventionsphase, bestehend jeweils aus mehreren Messungen. Zum Single-subject research design gehören aber noch weitere Arten als das AB Design, das:

- Withdrawal Design (ABA Design), welches zusätzlich zum AB Design mit einer Baseline abschließt,
- Multiple Baseline Design, welches mit der Dauer der Baselines bei mehreren Probanden, in unterschiedlichen Settings oder bei verschiedenen Aufgabenstellungen »spielt«,
- Alternating Treatments Design, welches die relative Wirksamkeit mindestens zweier Interventionen bei einem Patienten untersucht.

Möchte ein Therapeut das Single-subject research design beim eigenen Patienten anwenden, so sollte er beachten, dass es im Vergleich zu den systematischen Beobachtungen (► Buch: Kap. 3) mit einem größeren Aufwand verbunden ist. Die Evidenz ist dafür höher.

bzw. Meta-Analysen die höchste Evidenz. Die Validität ökonomischer Studien lässt sich mithilfe von Leitfragen überprüfen, z. B. ob es Angaben darüber gibt, aus welcher Perspektive Kosten und Nutzen analysiert wurden und ob das Follow-up ausreichend lang war. Wie deutlich sich die Ergebnisse der Interventionen unterscheiden, lässt sich anhand verschiedener ökonomischer Kenngrößen, z. B. des Kosten-Effektivitäts-Quotienten oder der ICER (Incremental Cost-Effectiveness Ratio), beurteilen. Ergänzend dazu gibt es gängige, aber nicht verbindliche Entscheidungshilfen, z. B. ein Bereich zwischen 20.000–30.000 €/QALY, um eine Intervention zu bevorzugen. Anhand von Leitfragen überprüft der Therapeut noch weitere Aspekte, wie bedeutsam die Ergebnisse sind, z. B. wie stark sich die Ergebnisse bzw. Schlussfolgerungen aufgrund der Sensitivitätsanalyse ändern.

Am Schluss beurteilt der Therapeut anhand von Leitfragen, ob er die Evidenz bei seinem Patienten anwenden kann. Dazu schätzt er ab, ob der eigene Patient einen gleichartigen gesundheitlichen Nutzen und gleiche Kosten erwarten kann. Zudem berücksichtigt er die Wertvorstellungen und Präferenzen des Patienten und beurteilt, ob er die Behandlung auf dieselbe Art und mit denselben Geräten und Materialien wie in der Studie durchführen kann.

Arten wissenschaftlicher Publikationen

Zusammenfassung

Wissenschaftlern und Wissenschaftlerinnen stehen verschiedene Studiendesigns zur Verfügung, um Forschungsfragen zu beantworten.

Primärstudien präsentieren systematisch gesammelte Originaldaten von Patienten. Für die wissenschaftliche Qualität spielen 2 Faktoren eine wichtige Rolle: die Kontrollgruppe und die Anzahl untersuchter Patienten. Zudem erhöht die zufällige Zuteilung der Patienten zur Interventions- oder Kontrollgruppe das Evidenzniveau. Daher ist die randomisiert kontrollierte Studie die wissenschaftlich wertvollste Primärstudie, um die Wirksamkeit einer Intervention nachzuweisen.

Integrative Publikationen fassen die Ergebnisse der Primärstudien und evtl. anderer integrativer Publikationen zusammen. Sofern sie systematisch durchgeführt wurden, stehen sie auf wissenschaftlich hohem Niveau. Für Therapeuten sind die systematischen Übersichtsartikel, Meta-Analysen und Leitlinien besonders relevant.

Kritische Bewertung gesundheitsökonomischer Studien

Zusammenfassung

Ökonomische Studien dienen dazu, die Kosten und den gesundheitlichen Nutzen verschiedener medizinischer oder therapeutischer Interventionen zu vergleichen, um herauszufinden, ob sich eine Intervention aus ökonomischer Sicht lohnt bzw. welche Intervention aus alternativen Vorschlägen zu bevorzugen ist.

Es gibt 3 wichtige Formen gesundheitsökonomischer Analysen, welche sich durch die Art des Nutzens unterscheiden: Bei der Kosten-Effektivitäts-Analyse besteht der Patientennutzen z. B. in der Anzahl gewonnener Lebensjahre, schmerzfreier Tage oder Verbesserung im Selbsthilfestatus, bei der Kosten-Nutzwert-Analyse in der Anzahl QALYs (qualitätsangepasste Lebensjahre) und bei der Kosten-Nutzen-Analyse im Geldbetrag, welcher durch die Anwendung der Intervention eingespart wird.

Die Studiendesigns und deren Hierarchien entsprechen denjenigen der ► Buch: Kap. 9–20. Bei z. B. Wirksamkeitsstudien bringen randomisiert kontrollierte Studien und systematische Übersichtsartikel

Sämtliche Übungsblätter (zu Kapitel 2, zu Kapitel 5, zu Kapitel 6) und das Excelfile »Berechnung der EPB-Größen« finden Sie unter <http://extras.springer.com>. Geben Sie im Suchfeld die ISBN 978-3-642-40635-5 ein.

Die Übungsblätter erleichtern Ihnen die praktische Anwendung der reflektierten Praxis, die Formulierung eigener »PICO-« Fragen und das gezielte Vorgehen bei der Literaturrecherche.

Das Excelfile hilft Ihnen, die im Buch beschriebenen Berechnungen der evidenzbasierten Praxis ganz einfach durchzuführen und zu interpretieren. Sie müssen nur wenige Daten, die Sie aus einem wissenschaftlichen Artikel entnehmen, in deutlich markierte Felder eintragen.