

# Designs für Einzelfälle und kleine Stichproben

## 2.1 Untersuchungsformen – 8

### 2.1.1 Interventionsstudien – 8

### 2.1.2 Diagnostische Untersuchungen – 9

### 2.1.3 Fallstudien – 11

### 2.1.4 Feststellung der Wirksamkeit einer Intervention – 13

## 2.2 Klinisch-diagnostische Designs – 14

### 2.2.1 Designs ohne Kontrollgruppe – 14

### 2.2.2 Kontrollgruppendesigns – 15

### 2.2.3 Messwiederholungsdesigns – 18

### 2.2.4 Phasendesigns – 19

## Literatur – 24

Daten aus Studien mit kleinen Stichproben oder Einzelfällen erfordern nicht nur eine besondere statistische Behandlung, ihre Genese ist zudem auch an besondere Untersuchungsdesigns gebunden, mittels derer die Verlässlichkeit und Fehlerfreiheit der Daten sichergestellt werden soll. Eine Unterscheidung wird dabei zwischen Designs mit und ohne Kontrolle vorgenommen, da Kontrollbedingungen insbesondere zur Sicherstellung der internen Validität dienen.

## 2.1 Untersuchungsformen

Allgemein werden im klinisch-diagnostischen Kontext drei Untersuchungsformen unterschieden:

- *Interventionsstudien.* Hiermit soll insbesondere die Wirkung therapeutischer Maßnahmen bei Patienten nachgewiesen werden. Typisch sind einzelfallanalytische oder stichprobenbezogene Untersuchungsdesigns, die die interne Validität sicherstellen sollen.
- *Diagnostische Untersuchungen.* Hier stehen die statistische Überprüfung diagnostischer Hypothesen und deren zufallskritische Absicherung im Vordergrund. Zur Beurteilung entsprechender Befunde dienen zumeist Testnormen, die aus Referenzstichproben gewonnen wurden. Dabei ist entscheidend, dass für die verwendeten Testverfahren zuverlässige Angaben zur Reliabilität vorliegen.
- *Fallstudien.* Fallstudien erstrecken sich von explorativen Fallbeschreibungen (*case studies*) bis hin zu experimentellen Einzelfallstudien (*single-case experimental studies*).

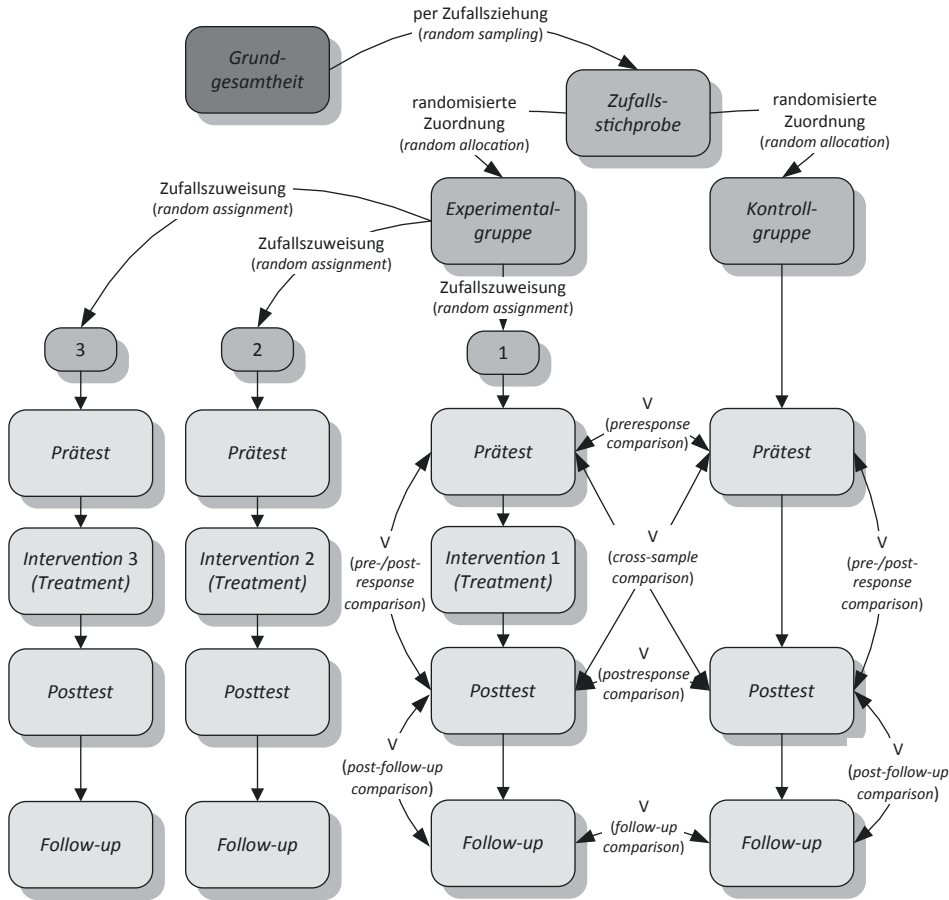
### 2.1.1 Interventionsstudien

Interventionsstudien finden im klinischen und medizinischen Kontext statt und verwenden *randomisierte Kontrollstudien* (*randomized controlled trial*). Diese gelten als „Goldstandard“ der klinischen Forschungsdesigns und liefern Ergebnisse für die *evidenzbasierte Forschung* (was übersetzt so viel wie „nachweisorientierte Forschung“ bedeutet). Damit verbindet sich der Anspruch (ursprünglich kommend aus der Medizin), kranken- bzw. patientenorientierte Entscheidungen nach Möglichkeit auf der Grundlage von klinischer Expertise (Fachwissen) und empirisch nachgewiesenen Wirksamkeitsstudien (externer Evidenz) zu treffen (Guyatt et al. 1992). Auch sollen Kontrollgruppendesigns davor schützen, den Erfolg einer Intervention von einer möglichen *Spontanremission* (einer eintretenden Besserung ohne jegliche Intervention) zu differenzieren (■ Abb. 2.1); wir werden uns im Folgenden noch mit solchen Designs beschäftigen.

Das Konzept evidenzbasierter Forschung ist allerdings nicht unumstritten, da insbesondere im Bereich der Psychotherapie durch die intensivere Erforschung bestimmter Therapieverfahren der Eindruck entstehen kann, dass die im Kontext von Forschung bevorzugten Therapiemethoden valider und effektiver seien als die, zu denen eine geringere Zahl empirischer Studien vorliegen. In der Folge kann dies zu einer Verdrängung und einem Ausschluss von (durchaus effektiven) Verfahren führen, die der Definition einer evidenzbasierten Methode nicht genügen.

Für den klinischen Forschungsprozess wird allgemein ein hierarchisches Phasenmodell angenommen, in dem neben explorativen Studien, randomisierten Kontrollstudien, Metaanalysen und Kosten-Nutzen-Analysen auch Einzelfall- und Kleingruppendesigns eine wichtige Rolle spielen (■ Abb. 2.2). Das damit einhergehende methodische Vorgehen, über die Identifikation einer vielversprechenden Idee, über klein- oder dann größer skalierte Studien zu Replikationen und schließlich finalen Kosten-Nutzen-Abschätzungen zu gelangen, wird auch in anderen Wissenschaftsbereichen angewendet.

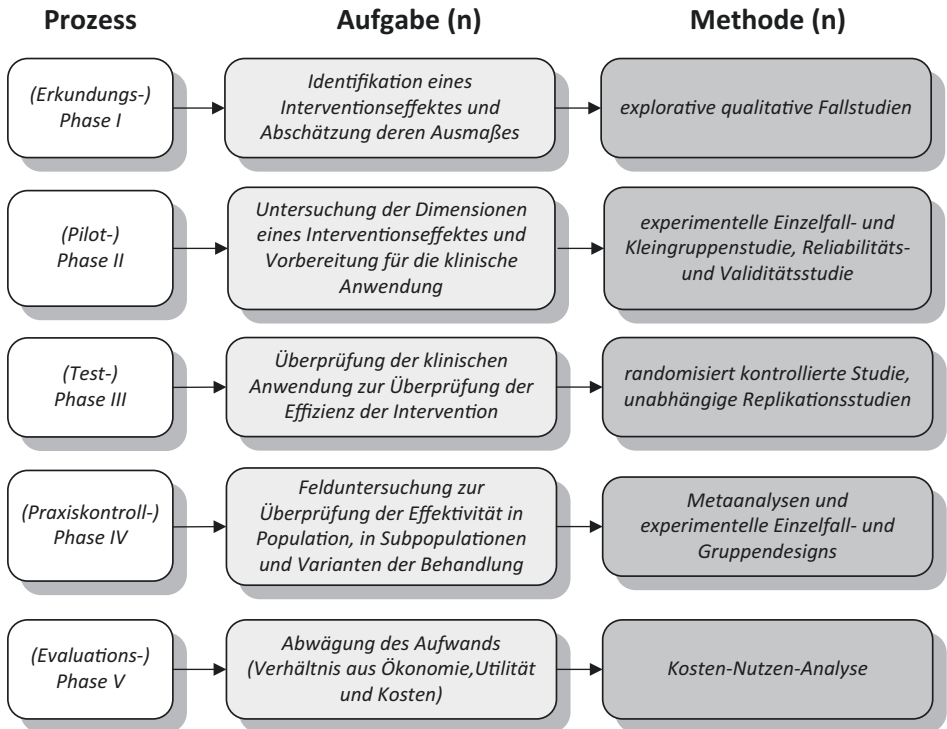
## 2.1 · Untersuchungsformen



■ **Abb. 2.1** Beispielhafte Darstellung einer randomisierten Kontrollstudie, hier mit einer Kontrollgruppe und drei Experimentalgruppen; zwischen der ersten Experimental- und der Kontrollgruppe sind verschiedene Vergleiche (V) eingezeichnet

### 2.1.2 Diagnostische Untersuchungen

Diagnostische Einzelfalluntersuchungen dienen dazu, die aus einer wissenschaftlichen Fragestellung abgeleiteten diagnostischen Hypothesen zu überprüfen und nach Möglichkeit zufalls-kritisch abzusichern (■ **Abb. 1.3**). Ausgangspunkt ist dabei ein Problem, ein Hilfesuch oder eine generelle Frage, die durch einen Klienten selbst, eine andere (beauftragte oder verantwortliche) Person oder eine Institution an den Diagnostiker herangetragen wird (vgl. Pospeschill und Spinath 2009). Obwohl auch hier die Formulierung einer Untersuchungshypothese und die Umsetzung in eine adäquate Untersuchungsform zentral ist, spielt doch seltener die systematische Variation von Untersuchungsbedingungen eine entscheidende Rolle als vielmehr die Erfassung einer Merkmalsausprägung eines Probanden unter möglichst konstanten (nonmanipulativen) Bedingungen mithilfe einer objektiven Methode (z. B. eines psychometrischen Tests). Die Beurteilung der Daten erfolgt dann auch nicht über den Vergleich mit Kontrollwerten, sondern



■ **Abb. 2.2** Prozessmodell des klinisch-diagnostischen Forschungsprozesses. (Nach Robey 2004; übersetzt und verändert)

durch Bezug auf eine adäquate und möglichst aktuelle Testnorm (Standardnorm, die aus repräsentativen Stichproben gewonnen wurde).

Dabei dient die diagnostische Untersuchung keinem Selbstzweck, sondern soll Schritte einer weiteren Behandlung vorbereiten. Entsprechend mündet der diagnostische Prozess in einem diagnostischen Urteil (zumeist in Form eines Gutachtens), in dem spezifische Empfehlungen und/oder Entscheidungsvorschläge zur angestrebten Intervention gegeben werden.

Auch hier spielt die Validität von Diagnosen eine entscheidende Rolle, unabhängig davon, ob eher klinische (informelle) oder statistische (formelle) Methoden der Datenintegration zum Einsatz kommen. Dabei gilt eine statistische Urteilsbildung allgemein als überlegen, da Verzerrungs- und Fehlertendenzen besser kontrolliert werden können, sofern die allgemein anerkannten Entscheidungsregeln konsequent angewendet werden.

Natürlich besitzen aber auch statistische Urteilsmodelle ihre Grenzen, da ihre Qualität wesentlich von der Absicherung der verwendeten diagnostischen Instrumente abhängt, z. B. von den Haupt- und Nebengütekriterien (vgl. Pospeschill 2010). Auch bleiben insbesondere Effekte, die sich nicht direkt mit empirischen Mitteln erfassen lassen, unberücksichtigt, wie kontextuelle Faktoren oder Effekte und Veränderungen, die außerhalb des Messverfahrens (oder generell außerhalb einer sicheren Messbarkeit) liegen. Gelegentlich verleiten statistische Aussagen auch zu unangemessenen Generalisierungen und nicht statthaften Kausalattributionen. Dabei wird nicht selten der Wahrscheinlichkeitscharakter statistischer Aussagen verkannt und dass statistische Vorhersagen nicht automatisch Verursachung bedeuten.

Dennoch finden formelle Methoden bei der Anwendung diagnostischer Fragestellungen breite Anwendung. Diagnostische Untersuchungen werden dabei nach den Methoden wissenschaftlicher Experimente durchgeführt, also mit expliziten Hypothesen, geplanten und kontrollierten Untersuchungsbedingungen sowie standardisierten Analyseverfahren. Für den diagnostischen Prozess bedeutet das auch, dass Entscheidungsregeln expliziert, Prognosegenauigkeiten sowie Angaben zur Validität einer Diagnose angegeben werden.

### 2.1.3 Fallstudien

In den ersten Phasen eines klinisch-diagnostischen Forschungsprozesses sind Einzelfallstudien unerlässlich, um erste Einblicke in die Wirkungsweise eines neuen Interventionskonzeptes zu gewinnen. Sie dienen hier insbesondere der Hypothesengenerierung hinsichtlich spezifischer Aspekte. Methodisch kann es sich dabei um ein (quantitatives) *Einzelfallexperiment*, um eine (quantitative) nichtexperimentelle *Einzelfallanalyse* oder um eine (deskriptive, qualitative) *Fallstudie* handeln (vgl. Kazdin 1982; Kratochwill 1978; Kratochwill und Levin 1992).

Bei Einzelfallexperimenten und der Einzelfallanalyse spielt vor allem die Schaffung von Kontrollbedingungen eine zentrale Rolle. Da hier eine Kontrollbedingung nicht durch eine zusätzliche (Kontroll-)Gruppe geschaffen werden kann, müssen entsprechende zufallsgesteuerte Bedingungsvariationen innerhalb desselben Individuums erfolgen. Nur durch adäquate Einzelfalluntersuchungsdesigns kann dabei der Forderung nach interner Validität (also dass beispielsweise eine Veränderung möglichst eindeutig – „kausal“ – auf den Effekt einer Intervention zurückgeführt werden kann) Rechnung getragen werden, z. B. durch ein *Multiple-Baseline-Design*. Dadurch gewonnene statistische Parameter auf der Ebene von Untersuchungseinheiten können schließlich auch zu Stichprobenaussagen zusammengefasst (agglutiniert) werden; dies erweitert den Aussagegehalt. Da dadurch aber keine repräsentativen Stichproben entstehen, bleibt ein Rückschluss auf andere Individuen eingeschränkt; hier besteht also ein Induktionsproblem. Auch lassen sich einzelne Individuen, die als separate Grundgesamtheit konzipiert wurden, erst dann zu einer Gruppe zusammenfassen, wenn die problemrelevanten Parameter der einzelnen Grundgesamtheiten hinreichend bekannt sind. Die Frage, ob Befunde aus Einzelfällen auch auf andere Probanden übertragbar sind, richtet sich daher nach dem Verhältnis intra- vs. interindividueller Varianz der relevanten Merkmale, nach dem Ausmaß der Vernachlässigbarkeit differenzieller Effekte und nach dem Grad der Replizierbarkeit.

Insbesondere Einzelfallstudien werden verwendet, um den Effekt einer Intervention bei einem einzelnen Fall zu überprüfen. Um solche Effekte zufallskritisch abzusichern, werden entsprechende Einzelfalldesigns notwendig. Erkenntnisse aus diesem Vorgehen können ebenso dazu verwendet werden, eine Interventionsform so zu adaptieren, dass ein möglichst maximaler Effekt entsteht. Dabei bieten sich Einzelfallstudien unter bestimmten Bedingungen besonders an:

- Wenn die Beobachtungsmöglichkeiten eines bestimmten Merkmals begrenzt oder schwer zugänglich sind bzw. wenn sich entsprechende Ereignisse zeitlich und räumlich spärlich verteilen (z. B. selten auftretende Erkrankungen, spezifische neurologische Schädigungen).
- Wenn es nicht gelingt, mit einem vertretbaren Aufwand eine hinreichend homogene und angemessen große Stichprobe von Merkmalsträgern zu akquirieren (wenn z. B. klinische Stichproben hinsichtlich Alter, Geschlecht, Intelligenz, Dauer des Klinikaufenthalts, sozioökonomischer Status, physiologischer Faktoren oder relevanter Persönlichkeitsmerkmale gematcht werden müssen, sinkt die Patientenzahl erheblich). Wird aber versucht, die Stichprobe zu vergrößern, so nimmt die Heterogenität der Stichprobe unverhältnismäßig

zu, und inferenzstatistische Probleme sind zwangsläufig die Folge (zumal unklar bleibt, auf welche Grundgesamtheit sich die späteren Ergebnisse übertragen lassen sollen, wenn die Patientenmerkmale weit streuen).

- Wenn die Unabhängigkeit interindividueller Beobachtungswerte nicht als gegeben angenommen werden kann (z. B. bei Untersuchung einer lokalen Patientengruppe auf einer Klinikstation oder in einer geschlossenen Klinikabteilung, sodass eine gegenseitige Beeinflussung von Patienten nicht ausgeschlossen werden kann).
- Wenn die Genese einer Patientenauswahl nicht mehr als Zufallsstichprobe angesehen werden kann. Dieser Fall kann eintreten, wenn Patienten über einen längeren Zeitraum (ggf. über mehrere Jahre) akquiriert werden müssen, um ausreichend viele Daten zu generieren. Ändern sich in dieser Zeit aber die Ein- oder Ausschlusskriterien, um zur Stichprobe zu gehören, die Indikatoren der Symptomatik, das Behandlungsverfahren etc., entstehen daraus ggf. neue Stichproben. Typischerweise arbeiten klinisch-psychologische Studien mit anfallenden Stichproben aus Patienten, die freiwillig teilnehmen oder durch ein Honorar zur Teilnahme „motiviert“ werden. Welche Selektionsmechanismen hierbei wirken, bleibt häufig unklar.
- Wenn langwierige Untersuchungen über einen längeren Zeitraum notwendig sind (z. B. Langzeitexperimente zu einem speziellen Training oder einer Rehabilitation oder Follow-up-Erhebungen zu späteren Zeitpunkten).
- Wenn ethische Kriterien die sofortige Anwendung an einer Stichprobe verbietet (z. B. wenn der Einsatz unerprobter psychotroper Substanzen erstmals getestet wird, unerwünschte Nebenwirkungen aber noch unbekannt sind und die optimale Dosierung noch beobachtet werden muss).

Ist ein Effekt bekannt, kann es notwendig sein, diesen unter variierenden Bedingungen oder Umständen zu untersuchen (z. B. im Rahmen einer systematischen Replikation), um seine Stabilität zu bestätigen; spezifische Merkmale einzelner Patienten machen eine Anpassung von Interventionen dabei häufig ohnehin notwendig. Dies kann dann eine diagnostische oder therapeutische Studie (bei der die Hypothesen vorwiegend intraindividuell definiert sind) in ein Einzelfall- oder Forschungsexperiment überführen, bei dem die im Einzelfall gefundenen Gesetzmäßigkeiten im Vordergrund stehen. Diese strengste Form der Messung psychologischer Größen bei einem Individuum erfordert, dass ein Phänomen durch systematische Variation der Bedingungen untersucht, angemessen statistisch ausgewertet und schließlich zufallskritisch abgesichert wird. Eine experimentelle Kontrolle psychischer Phänomene gelingt dabei allerdings nur, wenn die Variationsursachen und ihre Auswirkungen auf die symptomrelevanten Variablen bekannt und vorhersagbar sind.

Messtechnisch entscheidend ist, dass aufgrund der intraindividuellen Variabilität empirischer Daten (z. B. bedingt durch die Tagesform eines Patienten) eine statistische Verteilung der Messwerte für einen Einzelfall angenommen werden muss, die bei Anwendung statistischer Inferenzverfahren und beim Untersuchungsdesign zu berücksichtigen ist. Treffen bestimmte Verteilungsannahmen nicht zu, können klassische parametrische Tests nicht eingesetzt werden. Zudem können wiederholte intraindividuelle Beobachtungswerte von stochastischen Abhängigkeiten betroffen sein, die bei der Analyse der Daten berücksichtigt werden müssen. Datenanalyseverfahren, die eine Unabhängigkeit der Beobachtungsdaten erwarten, sind hier ebenfalls ungeeignet.

Unabhängig von diesen methodischen Implikationen interessiert sich aber ein Therapeut oder Diagnostiker weniger für Mittelwertsunterschiede zwischen Klientengruppen als vielmehr dafür, bei welchem Patienten eine Intervention eine signifikante Veränderung herbeiführt und bei welchem Klienten eine therapeutische Maßnahme ohne Wirkung bleibt. Individuelle Effekte können allerdings mit gruppenspezifischen Versuchen nicht entdeckt werden.

### 2.1.4 Feststellung der Wirksamkeit einer Intervention

Bei der Überprüfung von Interventionseffekten spielt generell die *Effektivität* oder *Wirksamkeit* hinsichtlich definierter Ziele oder Vergleichsgrößen (Zielzustand) eine zentrale Rolle. Dabei gibt es zwischen Ausgangs- und Zielzustand zumeist aber keinen normierten Mindestabstand (Effektgröße). Das Konzept der Wirksamkeit bedarf daher in aller Regel einer Konkretisierung und Präzisierung und ggf. auch einer eindeutigen Operationalisierung. Einigkeit besteht zumeist nur darin, dass sich der Effekt einer klinischen Intervention multimodal, d. h. auf verschiedenen Ebenen, bezüglich verschiedener Konstrukte und unterschiedlicher Datenebenen einstellt, was allerdings eine zusammenfassende Evaluation der Wirksamkeit erschwert. Die Beurteilung einer induzierten Veränderung kann somit verschiedenen Kriterien folgen:

- *Statistische Signifikanz als Maß der Veränderung.* Bei Gruppenstudien (im Vergleich zu einer Kontrollgruppe) und Einzelfall- oder Kleingruppenstudien (bei Prä-post- bzw. Baseline-Treatment-Vergleichen oder Vergleichen mit einer Testnorm) kann die statistische Signifikanz als Veränderungskriterium dienen, um einen Effekt gegen den Zufall abzusichern. Auch dient das Signifikanzmaß im diagnostischen Kontext dazu, für das erzielte Testergebnis ein Vertrauensintervall zu bestimmen, das den wahren Wert einschließt.
- *Praktische Signifikanz als Maß einer Veränderung.* Aus klinischen Kontrollstudien oder Eichstichproben können Effektgrößen (eine Mindestdifferenz bei der Veränderungsmessung) oder Normbereiche (zu Personen ohne – statistisch – auffälligen Befund) definiert werden. Dabei sind allerdings ausreichend große und homogene Stichproben sowie hinreichend reliable Untersuchungsinstrumente vorauszusetzen.
- *Klinische Signifikanz als Maß der Veränderung.* Bei der Beurteilung der klinischen Signifikanz geht es um die Einschätzung der klinischen Bedeutsamkeit einer Veränderung (z. B. auf der Verhaltensebene der betroffenen Person oder auch bei Effekten außerhalb der konkreten Intervention). Entsprechende Einschätzungen können von Experten oder Angehörigen von Patienten stammen, sind aber ggf. nur schwer zu validieren.
- *Prozentsatz wirksamer Interventionen oder agglutinierte Signifikanz einzelner Interventionen.* Denkbar sind hier die Festlegung einer Besserungsrate (Responder vs. Nonresponder bei einer Intervention) oder auch die statistische Abschätzung eines Overalleffektes über statistische Einzelbefunde. Dabei muss ggf. unterschieden werden, ob die Wirksamkeit über Kontrollstudien oder unter Feldbedingungen überprüft wurde. Entsprechend kann die Praxisbewährung ein eigenständiges Evaluationskriterium darstellen.
- *Wirkungsbreite und Wirkungsspektrum.* Die Möglichkeit verschiedener Wirkungen erfordert u. U. die Erfassung multimodaler Effekte, die über die eigentliche Zielsymptomatik hinausreichen. Dabei sind ggf. auch (erwünschte sowie unerwünschte) Nebeneffekte zu berücksichtigen, die nicht im Fokus der Intervention stehen. Diffuse Wirkungen können allerdings den Nachweis spezifischer Effekte erschweren bzw. deren Validität einschränken.
- *Veränderungsmuster.* Die Effekte einer Intervention können durch unterschiedliche Veränderungswerte (z. B. Prä-post-Vergleiche) betroffener Merkmale charakterisiert sein. Diese können insbesondere bei multimodalen Effekten auftreten. Dabei können auch Stadien der Veränderung eine Rolle spielen, die pro Zwischenziel eine Anpassung der Intervention erfordern.



- **Persistenz von Veränderungen.** Die Dauerhaftigkeit einer Veränderung kann durch Follow-up-Erhebungsphasen überprüft werden. Die Frage der Wirksamkeit über das Ende der Intervention hinaus steht dabei im Fokus. Anhaltende Effekte können dabei auch Teil eines Untersuchungsdesigns sein, wenn über den Zeitpunkt der eigentlichen Intervention hinaus weiter Daten erhoben werden.

Neben den Effektivitätskriterien einer Intervention, die den Zielabstand festschreiben, spielt natürlich auch die *Effizienz* eine wichtige Rolle, die den zur Zielerreichung notwendigen Aufwand (Verhältnis aus Kosten und Nutzen) beschreibt.

## 2.2 Klinisch-diagnostische Designs

Wird eine Interventionsmethode unter kontrollierten Bedingungen untersucht, so macht dies adäquate Versuchspläne erforderlich, die sich vor allem hinsichtlich des Vorhandenseins einer Kontrollgruppe unterscheiden. Untersuchungsdesigns mit Kontrollgruppe(n) sollen dabei vor allem einer Erhöhung der internen Validität dienen und damit eine möglichst eindeutige (kausale) Wirksamkeitszuschreibung gewährleisten.

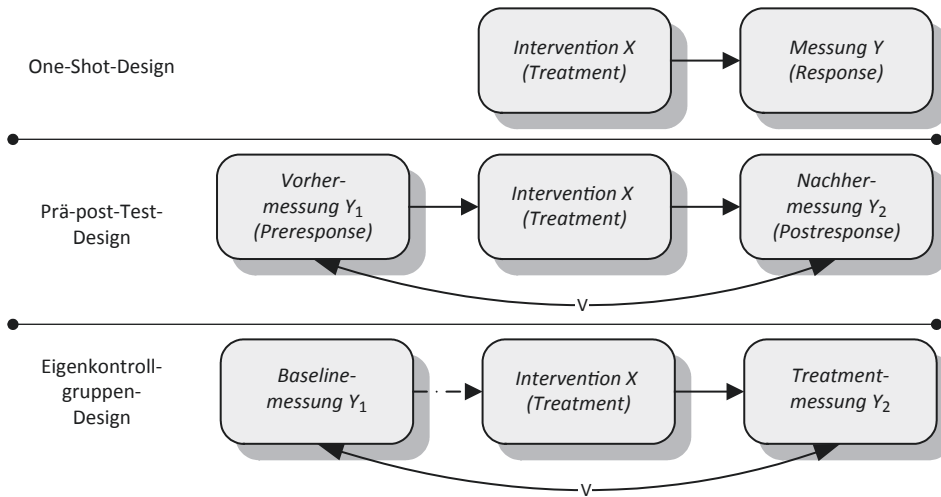
Generell wird unter einem *experimentellen Versuchsdesign* ein *Stichprobendesign* ( $n \geq 30$  pro Gruppe) unter Verwendung von *Erkenntnismethoden* – die im engeren Sinne der Beschreibung, Erklärung und Vorhersage von Verhalten und Erleben dienen sollen – verstanden. Unter einem *klinischen Versuchsdesign* wird hingegen zumeist eine *Einzelfallstudie* oder Stichprobenstudie sehr kleinen Umfangs ( $n < 15$ ) unter Verwendung von *Interventionsmethoden* – die der Prävention, Rehabilitation, Stabilisierung oder Veränderung von Verhalten und Erleben dienen sollen – verstanden. Die Bezeichnung „Versuchsdesign“ legt nahe, dass grundsätzlich spezifische wissenschaftliche Methoden zur Datengewinnung und Datenanalyse herangezogen werden.

### 2.2.1 Designs ohne Kontrollgruppe

Bei Designs ohne explizite Kontrollbedingung handelt es sich in der Regel um *Eingruppenpläne*, sodass die interne Validität nicht hinreichend gewährleistet werden kann (■ Abb. 2.3). Dazu misst das *One-Shot-Design* nur nach Abschluss der Intervention (Messung Y); es erhebt als Eingruppendesign retrospektiv Daten am Interventionsende. Das *Prä-post-Test-Design* ist ebenfalls als Eingruppendesign konzipierbar, erhebt allerdings zu zwei Zeitpunkten ( $Y_1$  und  $Y_2$ ) Daten und vergleicht die Messergebnisse (Veränderungsmessung V) vor und nach der Intervention (z. B. über eine Differenzenbildung der Messwerte). Das *Eigenkontrollgruppendesign* belässt die Stichprobe einige Zeit unbehandelt (Baselinemessung  $Y_1$ ), sodass eine künstliche Kontrollbedingung entsteht, bevor die Intervention einsetzt, und vergleicht die Messergebnisse zwischen diesen Phasen ( $Y_1$  und  $Y_2$ ). Da mögliche Zeiteffekte dabei unkontrolliert bleiben, ist auch hier die interne Validität eingeschränkt; zudem muss vorausgesetzt werden, dass die Baselinemessung keine Pretesteffekte auf die Intervention ausübt, da andernfalls beide Messungen nicht unabhängig voneinander wären.

Während in Erkundungsphasen (Phase I in ■ Abb. 2.2) oder auch zur Kontrolle praktischer Routinetätigkeit diese Designs durchaus sinnvoll sein können (z. B. um die generelle Wirkung einer Intervention abzuschätzen), werden für ausreichend reliable Abschätzungen Designs mit expliziten Kontrollbedingungen unerlässlich.





■ **Abb. 2.3** Eingruppenpläne ohne Kontrollbedingung mit Erhebungsphasen (Messung) und Interventionsphase (Treatment); das Kürzel „V“ steht für „Vergleich“

## 2.2.2 Kontrollgruppendesigns

Mit Beginn der Pilotphase (Phase II in ■ [Abb. 2.2](#)) sind Untersuchungsdesigns mit expliziten Kontrollbedingungen anzustreben, bei denen die Versuchspersonen randomisiert den Versuchsbedingungen zugeordnet werden (■ [Abb. 2.4](#)).

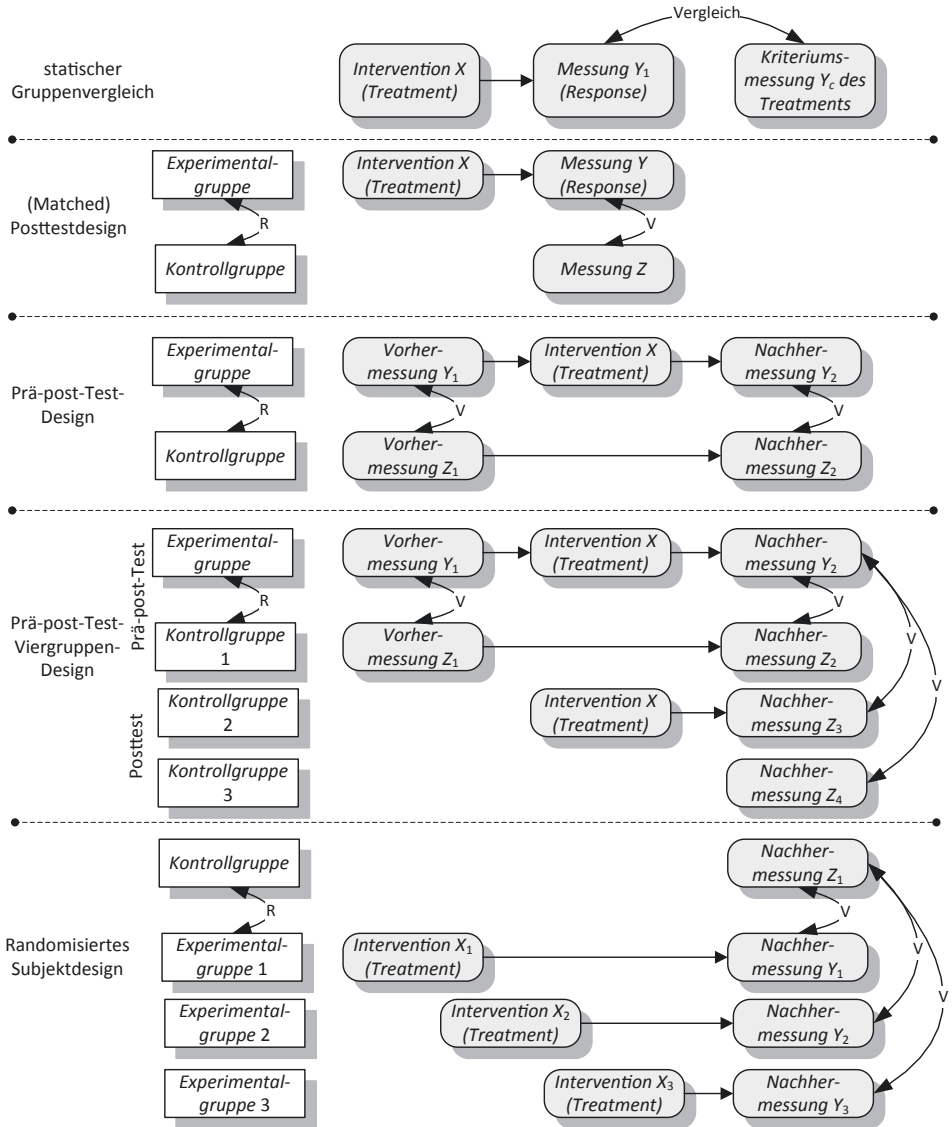
Dabei kann bei einem *statischen Gruppenvergleich* die Kontrollbedingung auch als externes Kriterium ( $Y_c$ ) herangezogen werden, das z. B. einer (statistischen) Norm oder einer vorab festgelegten Mindest- oder Schwellenwertdefinition entstammt (vgl. Pospeschill 2010).

Ein *Posttestdesign* verzichtet auf eine Vorhermessung und vergleicht eine (möglicherweise gematchte) Kontroll- und Experimentalgruppe anhand einer Einmalmessung (mit Intervention Y und ohne Intervention Z).

Das klassische *Prä-post-Test-Design* ergänzt das Posttestdesign durch eine Vorhermessung ( $Y_1$  und  $Z_1$ ) vor der Intervention. Der Vergleich erfolgt weiterhin zu einer Messung nach der Intervention ( $Y_2$  und  $Z_2$ ). Denkbar wäre hier auch, dass die Kontrollgruppe ein Placebo, eine Routinebehandlung oder eine andere Behandlungsform erhält.

Im klinischen Kontext kann eine Nichtbehandlung allerdings ethisch und klinisch problematisch sein und möglicherweise zu anderweitigen kompensatorischen Effekten (Spontanremission oder anderweitige Hilfe) führen. Auch sollte berücksichtigt werden, dass klinisch-psychologische Placebos keine vergleichbare Differenzierung zu einer Intervention erlauben (wie beispielsweise medikamentöse Placebos zu einer Wirksubstanz), da beide im Erleben und Verhalten ansetzen; dies gilt umso mehr, wenn Placebos zumindest Anzeichen oder sogar Elemente einer Intervention enthalten.

Sollen auch Kontrollgruppen mit in die Behandlung einbezogen werden, bieten sich Versuchspläne wie das *Prä-post-Test-Viergruppendesign* an, bei dem drei Kontrollgruppen in unterschiedlichen Variationen vor- und/oder nachgetestet werden, wobei eine Kontrollgruppe zusätzlich eine Intervention erhält. Damit werden alle Variationen einer Vorher-nachher-Messung (außer die der Experimentalgruppe) kontrolliert: nur eine Nachhermessung ( $Z_4$ ), eine Nachhermessung ( $Z_3$ ) mit vorgeschaltetem Treatment sowie eine Vorher-nachher-Messung ohne Treatment ( $Z_1$  und  $Z_2$ ).



■ **Abb. 2.4** Kontrollgruppenpläne mit Kontrollbedingung(en); das Kürzel „V“ steht für „Vergleich“, „R“ für „randomisierte Zuweisung“

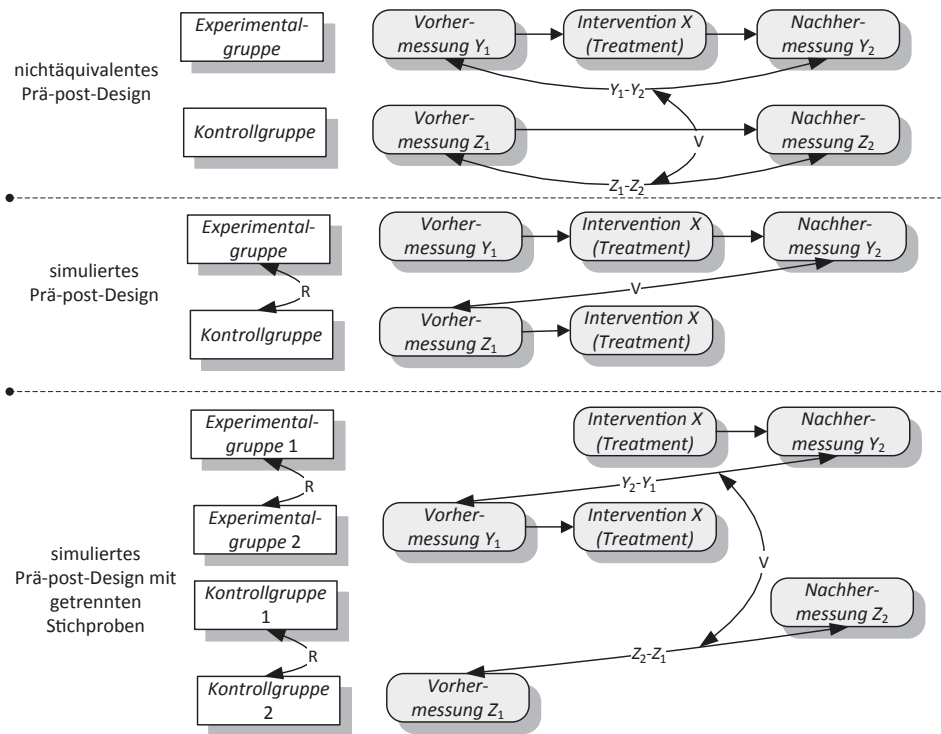
Alternativ kann auch das *randomisierte Subjekt-Design* verwendet werden, dass neben einer Kontrollgruppe ohne Behandlung mehrere Experimentalgruppen differenziert, die unterschiedlichen Interventionen (z. B. verschiedene Behandlungsalternativen oder eine Behandlung, die zeitversetzt appliziert werden) ausgesetzt werden. Bei den verschiedenen Interventionen (X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>) könnte es sich z. B. um andere Therapieformen oder um eine in ihren Parametern veränderte Behandlungsform handeln. Dabei wäre eine *Parametervariation* (z. B. eine Verlängerung oder Verkürzung der Sitzungsdauer oder der Anzahl der Sitzungen) oder eine *Parametermodifikation*

(z. B. die Hinzu- oder Wegnahme von weiteren Behandlungselementen) denkbar, deren Auswirkungen dann gemessen werden ( $Y_1, Y_2, Y_3$ ).

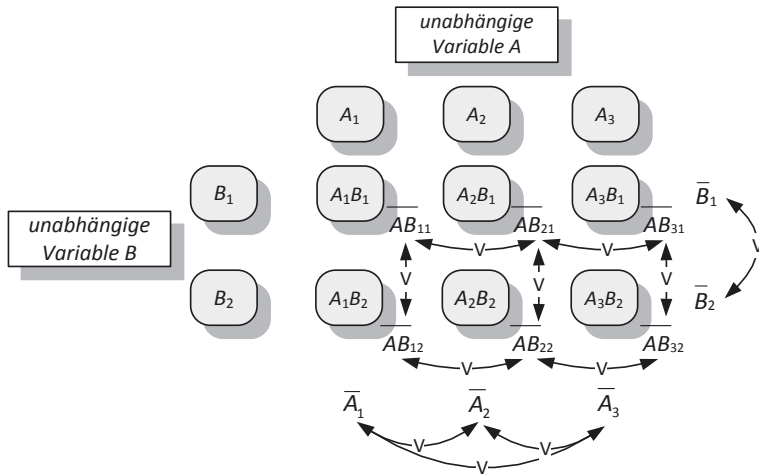
Während Eingruppen- (ohne Kontrollgruppe) und Zweigruppenpläne (mit Experimental- und Kontrollgruppe) mit Prä-post-Vergleich die Wirkung einer Intervention feststellen sollen, liefern Mehrgruppenpläne (mit mehr als einer Kontroll- oder Experimentalgruppe) einen Beitrag zur differenziellen Indikation, wenn es z. B. um die Kontrolle von Zeiteffekten (Mehrgruppen-designs) oder die Auswahl bzw. Optimierung von Interventionsformen (Subjekt designs) geht.

Weitere Varianten von Versuchsplänen mit einer Vorher- und Nachhermessung vergleichen mittels des *nichtäquivalenten Prä-post-Designs* die Differenzen ( $[Y_1 - Y_2] - [Z_1 - Z_2]$ ) aus den Wiederholungsmessungen zwischen Experimental- und Kontrollgruppe oder führen mittels des *simulierten Prä-post-Designs* einen Vergleich der Vorhermessung der Kontrollgruppe ( $Z_1$ ) mit der Nachhermessung der Experimentalgruppe ( $Y_2$ ) durch; dass hier auch für die Kontrollgruppe eine Intervention durchgeführt wird, bleibt unberücksichtigt. In einem komplexeren *simulierten Prä-post-Design mit getrennten Stichproben* wird dieser Vergleich mittels zweier getrennter Experimental- und Kontrollstichproben ( $[Y_2 - Y_1] - [Z_2 - Z_1]$ ) vorgenommen (■ Abb. 2.5).

Durch weitere Varianten und Kombinationen können schließlich komplexe (multifaktorielle und multimodale) Versuchspläne entstehen (■ Abb. 2.6). Dabei können als Faktor (A) nicht nur die Intervention (z. B. die Therapieform), sondern auch Merkmale der Interventoren (z. B. der Erfahrungsgrad von Therapeuten) und der Zielpersonen (z. B. Patientendiagnosen nach ICD – *International Statistical Classification of Diseases and Related Health Problems* oder



■ Abb. 2.5 Kontrollgruppenpläne mit Kontrollbedingung(en); das Kürzel „V“ steht für „Vergleich“, „R“ für „randomisierte Zuweisung“

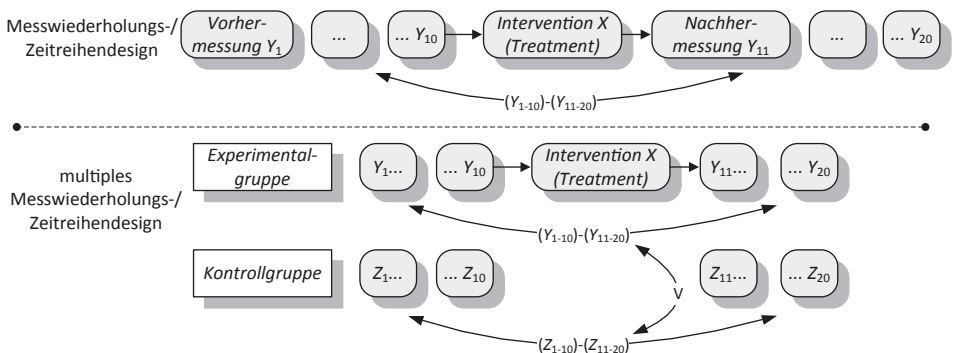


■ **Abb. 2.6** Beispiel für ein faktorielles Design mit zwei unabhängigen Variablen mit drei (z. B. verschiedene Patientendiagnosen) bzw. zwei (z. B. Erfahrungsgrad des Therapeuten hoch oder niedrig) Stufen. Durch die Kombination entstehen Spalten- (A), Zeilen- (B) und Zellenmittelwerte (AB), die spezifische Vergleiche erlauben; das Kürzel „V“ steht für „Vergleich“

DSM – *Diagnostic and Statistical Manual of Mental Disorders*) einbezogen (B) werden. Die so ermöglichte differenzielle Indikation kann auch komplexere Einflussstrukturen wie interagierende Faktoren ( $A_1B_1, A_2B_1, \dots, A_3B_2$ ) offenlegen. Ebenso können auf der Seite der Erfolgsmessung weitere (multivariate) Messkriterien definiert werden, die den Behandlungserfolg umfassender (und damit ggf. auch valider) erheben.

### 2.2.3 Messwiederholungsdesigns

Insbesondere für kleine bis sehr kleine Stichproben bzw. Einzelfallstudien bieten sich Messwiederholungspläne an, die sowohl vor als auch nach (ggf. auch während) einer Intervention (Mehrfach-)Messungen beinhalten (■ **Abb. 2.7**). Entsprechende *Messwiederholungs-* und *Zeitreihen-*designs sind dabei sowohl ohne Kontrollbedingung bzw. -gruppe als auch mit Kontrollgruppe



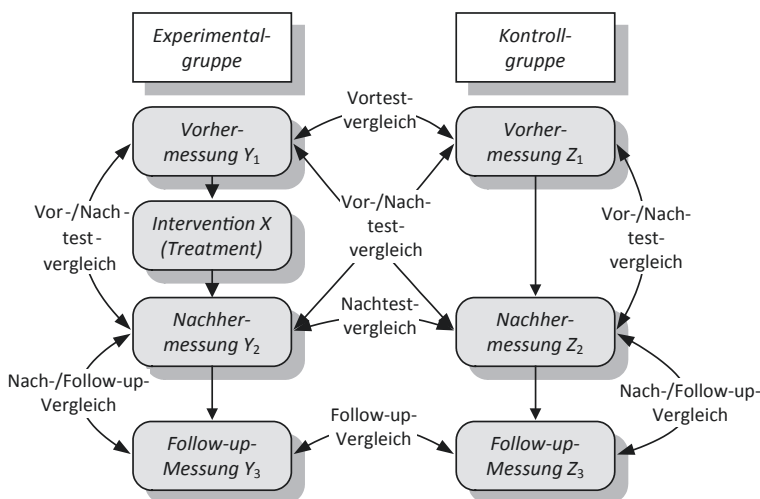
■ **Abb. 2.7** Messwiederholungspläne mit und ohne Kontrollbedingung; das Kürzel „V“ steht für „Vergleich“, „R“ für „randomisierte Zuweisung“

umsetzbar. Bei einer Zeitreihe (diese entsteht etwa ab 50 Messungen in Folge) wird dabei eine feinere Messauflösung in bestimmten Phasen (z. B. höhere Messraten zu Beginn und am Ende einer Intervention, um die Messgenauigkeit zu erhöhen) oder eine generell höhere Taktrate an Messwiederholungen (z. B. eine Verkürzung der Messintervalle, um zusätzliche Messzeitpunkte zu erzielen) angestrebt. Dies kann beispielsweise dann geschehen, wenn Erhebungen innerhalb einer Therapiesitzung (z. B. Erhebung der Redezeit oder von Interaktionssequenzen), zwischen den Sitzungen (z. B. Stimmungsbild eines Patienten vor, zwischen und nach einer Intervention) oder während eines vollständigen Klinikaufenthaltes (z. B. durch Führen eines Stimmungstagebuches, das morgens, mittags und abends vom Patienten die Beantwortung standardisierter Fragen erwartet) vorgenommen werden. Bei der Prozessanalyse solcher Zeitreihendaten wird nach Verlaufsmustern (z. B. zyklischen oder periodischen Mustern) gesucht und mit zeitgleichen Ereignissen oder Vorkommnissen in Verbindung gebracht (*event related correlation*).

Insbesondere für differenzielle Studien kann entscheidend sein, längerfristige Effekte im Design zu berücksichtigen. Damit wird es notwendig, neben dem Interventionsbeginn und dem Interventionsende ggf. einen dritten Messzeitpunkt, der einer Nacherhebung dient (*follow-up*), mit aufzunehmen (■ Abb. 2.8). Typisch sind solche Designs u. a. bei der Prozessforschung, die Langzeiteffekte (als Behandlungserfolg) über das Ende einer Intervention hinaus untersucht. Bei einer solchen katamnestischen Untersuchung (als Gegenstück zu einer anamnestischen Untersuchung oder kurz Anamnese) liegt die Follow-up-Phase häufig zeitlich deutlich später (z. B. ein Jahr), als die Postphase und kann zudem auch mehrere Messwiederholungen umfassen.

## 2.2.4 Phasendesigns

Besondere Untersuchungsdesigns sind notwendig, wenn es um die Analyse von Einzelfällen (*single case studies*) oder sehr kleinen Stichproben ( $n \geq 2$ ) geht. Diese orientieren sich vor allem an Messwiederholungs- und Zeitreihendesigns und werden als *Designs mit Phasenübergängen* oder kurz *Phasendesigns* bezeichnet. Bei den Messwiederholungen lösen sich dabei Erhebungsphasen



■ Abb. 2.8 Beispiel für ein Prä-post-Test-Follow-up-Kontrollgruppendesign und die Möglichkeiten empirischer Vergleiche zu den verschiedenen Messzeitpunkten

ohne Intervention bzw. Behandlung (Baseline- oder A-Phasen) von Phasen mit Intervention bzw. Behandlung (Treatment- oder B-Phasen) ab. In der Folge wird allgemein von AB-Designs gesprochen.

Bei Messwiederholungen (*repeated measurement*) ist es erforderlich, die verwendete(n) Messvorrichtung(en), die Messzeitpunkte (Anzahl, Taktung und Verteilungsart) und spezifische Kontextvariablen (weitere Umwelteinflüsse) zu standardisieren, um Artefakte, verzerrende und konfundierende Effekte ausschließen zu können. Die Anzahl der Messungen muss repräsentative Aussagen zulassen. Dabei kann die Messwiederholung eingeschränkt sein, wenn Wiedererkennungseffekte (bei zu kurz gewählten Zeitabständen), *Carry-over-Effekte* (Übertragungseffekten von Messung zu Messung) oder Ermüdungseffekte (ein Nachlassen von Leistung oder Konzentration aufgrund zu häufiger oder zu lang andauernder Messungen) nicht auszuschließen sind. In einem Phasendesign kann ein Effekt zudem nur dann festgestellt werden, wenn beim Übergang der Phasen nur eine Variable variiert wird, da ansonsten keine distinkte Erfassung des Effektausmaßes der UV möglich ist.

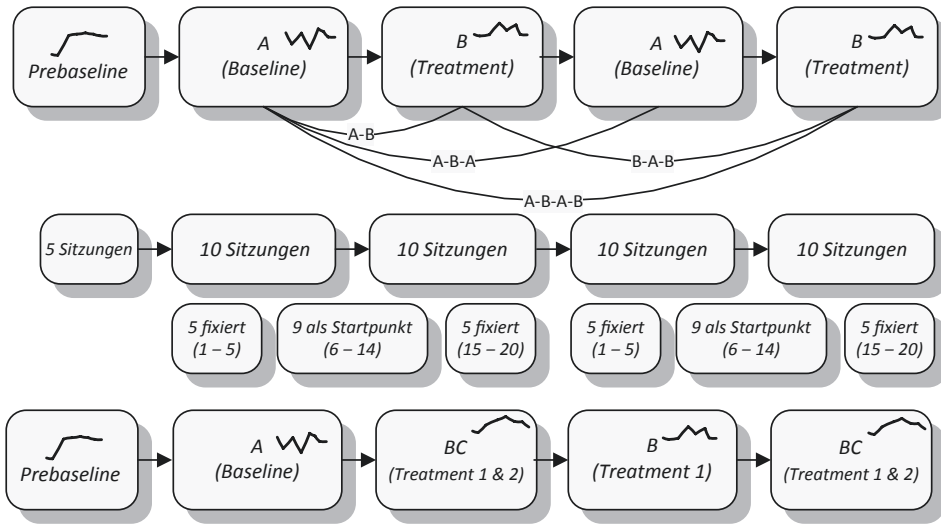
In der *Baselinemessung* der AV wird das Ausgangsverhalten (zumeist über mehrere Messungen, um Zufallsschwankungen auszugleichen) gegenüber den Treatmentphasen festgestellt. Für solche AB-Designs gilt daher, dass eine Treatmentphase nicht beginnen sollte, bevor sich die Baselinephase stabilisiert hat, d. h. keinen erkennbaren Trend mehr zeigt. Eine Ausdehnung der Baselinephase kann aber ethisch problematisch sein, z. B. im Zuge einer Verzögerung einer Medikamentengabe. Startet die Treatmentphase aber zu früh, bleibt es schwierig, eine Veränderung bedingt durch die Intervention eindeutig von einem generellen Trend (Trainings-, Übungs-, Anpassungs- oder Akkomodationseffekt) zu unterscheiden. Die Anzahl hierzu notwendiger Baselinesitzungen ist daher ggf. vorab festzulegen. Gleiches gilt für die minimale Anzahl der Treatmentsitzungen, die benötigt werden, damit sich die Wirkung der Intervention voll entfalten kann.

Ursachen für eine sich nicht einstellende Stabilität in der Baselinephase können vielfältig sein: eine spontane oder tagesformabhängige Verschlechterung oder Verbesserung im Krankheitsverlauf (Abwärts- oder Aufwärtstrend), ein Placeboeffekt (Abwärtstrend gefolgt von einem Aufwärtstrend) oder ein anhaltender Trainingseffekt (leichter Aufwärtstrend). Eine ausgiebige Baselinemessung kann zudem dazu dienen, anhand des gezeigten Datenmusters die Vergleichbarkeit zu anderen Patienten sicherzustellen.

Bei den Untersuchungsplänen für Einzelfälle werden *Withdrawal-/Reversal-Designs* (*withdrawal* bedeutet so viel wie „Absetzen“ oder „Entziehen“) von *Multiple-Baseline-Designs* unterscheiden. Die Gruppe der *Withdrawal-Designs* unterbricht mittels Variation eines A-B-A-Designs die Intervention. *Reversal-Designs* setzen dazu speziell eine Therapie mit umgekehrtem therapeutischem Ziel ein. Auch Kombinationen dieser Designtypen sind möglich.

Bei den *Withdrawal-Designs* werden unterschieden (■ Abb. 2.9):

- *A-B-Design*. Eine Baselinemessung gefolgt von einer Interventionsphase ist zwar eine zumeist praxisnahe Lösung (eingerahmt durch eine Anamnese und ggf. eine Nachbehandlungsphase), allerdings für die *Hypothesenprüfung* nur bedingt bis gar nicht geeignet. Es bleibt hier weitgehend unkontrolliert, welche Veränderungen auf den Krankheitsverlauf, die Intervention oder andere unkontrollierte Ereignisse zurückzuführen sind.
- *A-B-A-Design*. Ein Vergleich der zwei A-Phasen erlaubt eine Aussage über die Wirkungsdauer der Intervention. Zeigen die Daten mit Einführung der Intervention einen Aufwärtstrend (im Sinne einer Zunahme an Lebensqualität bzw. Abnahme der Symptomatik) und nach Absetzen einen Abwärtstrend, besteht eine gewisse Evidenz für die Wirkung des Treatments. Dies impliziert allerdings, dass ein Beenden der Intervention während der



■ **Abb. 2.9** Untersuchungspläne für Einzelfalldesigns

Datenerhebung angestrebt wird, da die Studie ggf. mit einer Verschlechterung im Befinden eines Patienten endet.

- **B-A-B-Design.** Ein Design, das mit einer B-Phase beginnt, ist zwar klinisch günstig (da es sofort mit der Intervention beginnt und mit der Intervention endet), allerdings eingeschränkt hinsichtlich der Frage der initialen Baseline (vor der Intervention), die hier unkontrolliert bleibt. Gegebenenfalls kann zum Zwecke einer Kontrolle eine Pre-Baseline-Phase vorgeschaltet werden, wenn anfängliche Trainingseffekte nicht auszuschließen sind.
- **A-B-A-B-Design.** Hiermit lassen sich die Konsequenzen eines Abbruchs bzw. Aussetzens einer Intervention und das erneute Einsetzen überprüfen. Dies kommt der Replikation eines Interventionseffektes gleich. Allerdings kann das Design in bestimmten Settings aus ethischen Gründen nicht umsetzbar sein, z. B. bei einer medizinischen Studie, bei der ein wirksames Medikament dadurch wieder abgesetzt werden müsste. Zur Steigerung der Evidenz bietet es sich hier an, neben den Zielvariablen auch Nichtzielvariablen mitzuerheben, da sich die Nichtzielvariablen je nach Abhängigkeitsgrad von den Zielvariablen mitverändern sollten, wenn die Intervention die Änderung der Zielvariablen verursacht hat.
- **A-BC-B-BC-Design/A-B-A-B-BC-Design.** Die aufgeführten Designs lassen sich vielfältig durch Kombination mehrerer Interventionen (hier B und C) variieren, die kombiniert und einzeln überprüft werden. Damit kann ein Interaktionseffekt geprüft sowie der Anteil der Kombinationswirkung von BC auf B zurückgeführt werden.

Weitere Varianten von *Withdrawal-Designs* sind:

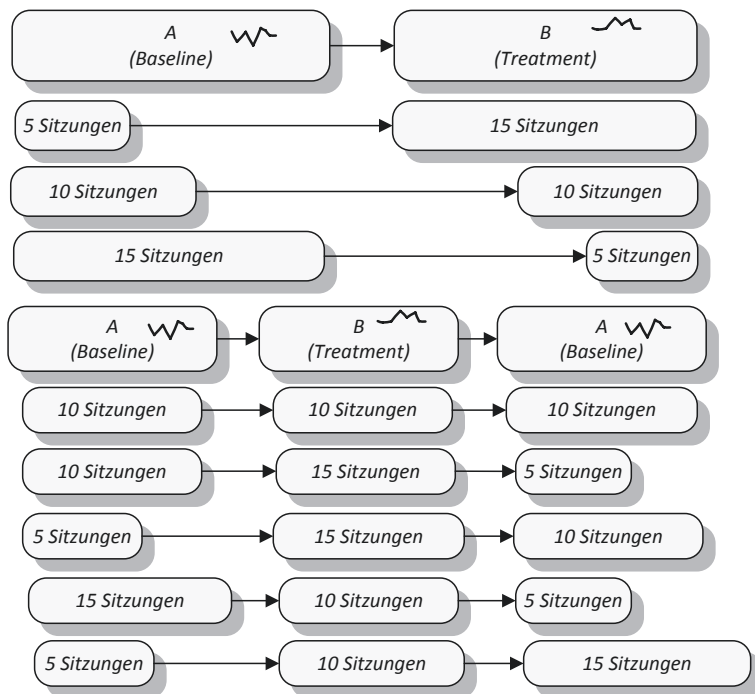
- das **A-B-A-B-A-B-Design**, das mehrfache Wiederholungen des A-B-Designs enthält,
- das **A-B-A-C-A-Design**, das zwei Interventionsmethoden mit der Baseline vergleicht, ohne diese allerdings zu kombinieren, oder
- das **A-B-A-B'-B''-B'''-Design**, das verschiedene Wirkungsintensitäten einer Intervention beinhaltet, z. B. eine stufenweise ausklingende Therapie.



- Ein besonderes Design ist schließlich das *Changing-Criterion-oder A-B-C-D-E-F-Design*. Hier wechseln sich verschiedene Interventionen ab, die – jeweils nach Erreichen einer vorab definierten Zielgröße – die Baseline für die nachfolgende Behandlungsphase darstellen, z. B. Zwischenziele als Therapie-/Trainingserfolg oder bestimmte Stufen einer Rehabilitation.

Bei den *Multiple-Baseline-Designs* handelt es sich um zeitversetzte Phasendesigns für eine oder mehrere Zielvariablen, die in die Behandlung einbezogen werden (■ Abb. 2.10). Dabei wird die Anzahl der Sitzungen systematisch variiert, d. h., die Gesamtzahl der Sitzungen erfolgt nach unterschiedlicher Aufteilung zwischen Baseline und Treatment. Diese Designs erlauben es, den Start des Treatments zwischen einzelnen Fällen zu staffeln (d. h. zu unterschiedlichen Zeitpunkten beginnen zu lassen) und damit Veränderungen eindeutiger auf das Treatment (gegenüber dem Zufall) zurückzuführen. Gleichzeitig impliziert die Betrachtung mehrerer Fälle einen Inferenzschluss im Hinblick auf die Wahrscheinlichkeit, dass die Wirkung des Treatments auf eine größere Grundgesamtheit übertragbar ist. Dabei werden folgende Varianten unterschieden:

- Multiple-Baseline-Design Across Behavior*. Dieses Design betrachtet einen Fall (Patienten) und mehrere Zielvariablen, für die zu Beginn simultan jeweils eine gesonderte Baseline erhoben wird. Die verschiedenen Interventionen beginnen zeitversetzt, sodass – unter der Annahme voneinander unabhängiger Zielvariablen – der zu beobachtende Effekt den jeweiligen Zielvariablen der Intervention zugeschrieben werden kann. Die Annahme der Unabhängigkeit der Zielvariablen setzt allerdings voraus, dass sich unter einer spezifischen



■ Abb. 2.10 Beispiele für Multiple-Baseline-Designs, bei denen die Anzahl der Sitzungen pro Phase systematisch variiert wird

Intervention auch ausschließlich die zugehörige Zielvariable ändert, während die anderen Zielvariablen unbeeinflusst bleiben.

- *Multiple-Baseline-Design Across Subjects*. Hier werden mehrere Fälle mit Bezug auf eine einzelne Zielvariable betrachtet. Dabei wird das Verhältnis von Baseline- und Treatment-sitzungen systematisch variiert, z. B. die Gabe eines Placebos in der Baseline und die Gabe eines Medikamentes in der Treatmentphase.
- *Multiple-Baseline-Design Across Settings*. Hier werden verschiedene Interventionssettings (z. B. verschiedene Intensitäten einer Therapie, eine unterschiedliche Behandlungsdauer oder verschiedene Kombitherapien) über verschiedene Fälle verglichen. Entsprechend werden mehrere Baselines erhoben. Ziel dabei ist es, das optimale Setting zu identifizieren oder die Settings zu vergleichen.
- *Nonconcurrent-Multiple-Baseline-Design Across Individuals*. Treten bestimmte Zielvariablen nur in seltenen Fällen auf, sodass nicht sofort mehrere Fälle zur Verfügung stehen, bietet sich dieses Design an. Dazu werden die Längen der Baselinephasen vorab definiert und den im Laufe der Zeit aufgenommenen Patienten per Randomisierung zugewiesen. Die interne Validität bleibt allerdings eingeschränkt, da durch den zeitlichen Versatz keine identischen Untersuchungsbedingungen garantiert werden können und ebenso die Frage der Stichprobendefinition Änderungen unterliegen kann.
- *Randomization-Design*. Das auch als *Alternating-Treatment* bezeichnete Design kombiniert zwei oder mehr Behandlungsmethoden durch ein schnelles Alternieren der Behandlungskomponenten, sodass die Untersuchungsbedingungen erhalten bleiben. Ziel ist es, die Wirksamkeit der Methoden zu vergleichen, wobei vorausgesetzt wird, dass sich die Behandlungskomponenten nicht wechselseitig beeinflussen. Das Design bei Einzelfällen angewendet erwartet aber zur Verallgemeinerung die Durchführung an mehreren Fällen. Auch wird durch eine Semirandomisierung eine Zufallsanordnung der verschiedenen Behandlungsmethoden vorausgesetzt, um Sequenzeffekte auszuschalten. Allerdings sind insbesondere *Carry-over-Effekte* nicht grundsätzlich auszuschließen. Durch unabhängige Prüfung der einzelnen Behandlungsmethoden in A-B-Designs kann das Ausmaß solcher Effekte aber überprüft werden.

*Multiple-Baseline-Designs* werden vor allem für Untersuchungen angewendet, bei denen nicht erwartet wird, dass die abhängige Variable nach Einsetzen des Treatments auf das Niveau der Baseline zurückfällt (sind darauf aber nicht beschränkt, vgl. ■ [Abb. 2.10](#)). Dies hat zur Folge, dass eine experimentelle Manipulation des Treatmenteffektes in der Regel nicht möglich ist. Hinzu kommt, dass ein Effekt (gegenüber den *Withdrawal-Designs*) nicht direkt gezeigt wird, sondern nur indirekt aus den Datenmustern der noch nicht behandelten Zielvariablen. Damit bleiben kausale Rückschlüsse ausgeschlossen.

In medizinischen Studien wird oft ein Vorgehen verwendet, das als *N-of-1 randomized controlled trial* bezeichnet wird. Dieses Vorgehen dient der Entscheidung über die Wirksamkeit einer Therapie (z. B. eines Medikaments), eines Patienten als Teil einer Studie, der Dosierung und ggf. den Nebenwirkungen. Dazu erhält der Patient Behandlungen, bei denen sich die zu überprüfende Therapie mit einer alternativen Methode (ggf. Placebo) in randomisierter Form abwechselt (häufig als Doppelblindversuch, bei denen Arzt und Patient nicht wissen, was aktuell appliziert wird). Die Veränderung in der Lebensqualität für den Patienten wird durch Tagebucheinträge erfasst. Für klinisch-diagnostische Anwendungen ist dieses Vorgehen allerdings nicht zu empfehlen. Lediglich für Fälle, in denen die Wahl des optimalen Treatments nicht geklärt ist, ist dieses Vorgehen möglich, wenn konkurrierende Methoden verglichen werden sollen.

Auf die Validität begünstigend wirken sich bei *Multiple-Baseline-Designs* *a priori* spezifizierte Hypothesen, ein vorab festgelegter Zeitrahmen und einheitliche Datenerhebungszeitpunkte aus.

Während bei *Multiple-Baseline-Designs* eine systematische Variation der Sitzungen zwischen Baseline und Treatment stattfindet, erfolgt bei *Withdrawal-Designs* der Übergang zwischen Baseline und Treatment in anderer Form (■ Abb. 2.9): Aus der Gesamtzahl der Sitzungen wird eine Mindestzahl an initialen Baseline- und finalen Treatmentsitzungen fixiert (z. B. 5 Sitzungen am Anfang und 5 Sitzungen am Ende einer Behandlung). Bei den verbleibenden Sitzungen (z. B. zwischen Sitzung 6 bis 14) entscheidet der Zufall, wann die Baselinesitzungen enden und das Treatment beginnt. Dies entspricht einer randomisierten Zufallszuweisung (*random assignment*) und stellt eine zufallsgesteuerte Bedingungsvariation dar, mit der die interne Validität sichergestellt werden soll.

## Literatur

---

- Guyatt, G., Cairns J., & Churchill D. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 268(17), 2420–2425.
- Kazdin, A.E. (1982). *Single – case research designs*. Oxford: Oxford University Press.
- Kratochwill, T.R. (Hrsg.). (1978). *Single subject research*. New York: Academic Press.
- Kratochwill, T.R., & Levin, J.R. (1992). *Single-case research design and analysis*. Mahwah, Erlbaum.
- Pospeschill, M., & Spinath, F.M. (2009). *Psychologische Diagnostik*. Stuttgart: Reinhardt (UTB).
- Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation*. Stuttgart: Reinhardt (UTB).
- Robey, R.R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders*, 37(5), 401–411.

Methoden für die klinische Forschung und  
diagnostische Praxis

Ein Praxisbuch für die Datenauswertung kleiner  
Stichproben

Pospeschill, M.; Siegel, R.

2018, XIII, 202 S. 147 Abb. Book + eBook., Softcover

ISBN: 978-3-662-54725-0