



Randomisiert. Kontrolliert. Doppelblind. Warum?

Peter Kleist

Quintessenz

- Die Aussagefähigkeit einer klinischen Studie hängt entscheidend von ihrer internen Validität ab. Eine hohe interne Validität entspricht einer bestmöglichen Vermeidung von studienimmanenten systematischen Fehlern (auch «Bias» genannt), die das Ergebnis der Studie verfälschen können.
- Ein Bias wird innerhalb von Studien unter anderem durch gezielte (und nicht auf Zufall basierende) Behandlungszuweisung, ein ungleiches Mass an Betreuung bzw. Supportivbehandlung der Studienpatienten sowie durch Suggestionseffekte erzeugt – letztere sowohl auf Seite der Patienten als auch auf der des Studienarztes.
- Die wichtigsten Instrumente, diese unterschiedlichen Formen eines Bias zu vermeiden, bestehen in der Einführung einer Kontrolle, der Randomisierung und der Verblindung der untersuchten medizinischen Massnahmen.
- Trotz ihren Limitationen ist die randomisierte, kontrollierte und doppelblinde Studie die verlässlichste und objektivste Methode mit der höchsten Aussagefähigkeit, weil sie am wenigsten für einen Bias anfällig ist.

Summary

Randomised. Controlled. Double-blind. Why?

- *The effectiveness of a clinical study depends decisively on its internal validity. High internal validity means avoiding, to the greatest extent possible, systematic errors inherent in studies (also called bias) which could falsify the study's results.*
- *Bias within studies arises from targeted (not chance-based) decisions about treatment, a variable degree of care or supportive treatment of the study population, and the effects of suggestion – the last-mentioned on the part of both patients and the doctor in charge of the study.*
- *The most important instruments for avoiding these various forms of bias consist in the introduction of control, randomisation and blinding of the medical measures under investigation.*
- *Despite its limitations the randomised, controlled, double-blind study is the most reliable and objective method with the highest degree of effectiveness, because it is the least liable to be affected by bias.*

Einführung

Selbst wenn Sie nicht aktiv an der Durchführung klinischer Studien beteiligt sind, werden Sie als Leser medizinischer Fachliteratur doch ständig mit ihnen konfrontiert. Dabei begegnen Sie immer wieder der sogenannten randomisierten, kontrollierten Doppelblindstudie. Dieses Studiendesign stellt nicht nur den «Goldstandard» innerhalb der modernen klinischen Arznei-

mittelentwicklung dar, sondern es ist auch mit dem höchsten Erkenntnisgrad im Rahmen der evidenzbasierten Medizin verbunden. Warum kommt der randomisierten, kontrollierten Doppelblindstudie eine dermassen hohe Bedeutung zu? Und was bedeuten «randomisiert», «kontrolliert» und «doppelblind» eigentlich? Der nachfolgende Artikel wird Ihnen diese Fragen beantworten.

Objektivität und Aussagekraft

In klinischen Studien werden Wirksamkeit, Verträglichkeit und Sicherheit medizinischer Massnahmen experimentell am Patienten untersucht. Doch von einem naturwissenschaftlichen Experiment weicht eine klinische Studie in einigen wesentlichen Punkten ab:

- Aufgrund seiner Einzigartigkeit reagiert jeder Patient individuell auf eine medizinische Massnahme – sowohl physisch als auch psychisch. Ob er auf eine Behandlung anspricht oder Nebenwirkungen auftreten, ist in der Regel nicht vorhersehbar. In einer klinischen Studie steht man daher grundsätzlich dem Problem der Inhomogenität von Stichproben gegenüber, und es ist mit unterschiedlichen individuellen Reaktionsmustern und Krankheitsverläufen zu rechnen.
- Vor allem bei Langzeitstudien lassen sich die äusseren Versuchsbedingungen nur teilweise standardisieren. Unterschiedliche Lebensbedingungen und Verhaltensweisen der Patienten sowie Begleiterkrankungen und -therapien können als «Störgrössen» das Ergebnis einer Untersuchung erheblich beeinflussen.
- Der Studienarzt ist kein aussenstehender, unbeteiligter Beobachter, sondern als behandelnder Arzt ein aktiver Bestandteil der Studie, der mit der «Beobachtungseinheit» Patient in eine Wechselwirkung tritt. Die Art der Patientenführung, unterstützende therapeutische Massnahmen sowie die Übertragung seiner eigenen positiven oder negativen Einstellung zur Behandlung auf den Patienten wirken sich auf das Ergebnis einer Studie aus.
- Viele Messungen unterliegen einem hohen Mass an Subjektivität. Das Schmerzempfinden eines Patienten wird zum Beispiel durch die Annahme beeinflusst, ein wirksames

Analgetikum erhalten zu haben. Aber auch der Arzt, der Befunde vor dem Hintergrund der von ihm durchgeführten Therapie erhebt und interpretiert, ist nicht frei von Subjektivität; glaubt er an die Wirksamkeit seiner Behandlung, kann sich das auf seine Einschätzung des klinischen Zustands eines Patienten oder die Interpretation von Befunden auswirken.

Da sich viele der genannten Faktoren nicht ausschalten lassen, kommt dem Umgang mit ihnen eine entscheidende Bedeutung zu. Ein gutes Studiendesign kann dafür sorgen, dass der Einfluss der obengenannten Störgrössen auf das Studienergebnis weitgehend reduziert wird. Die Aussagekraft und damit der Wert einer klinischen Studie besteht ja letztlich darin, dass nur die untersuchte Massnahme, zum Beispiel eine medikamentöse Behandlung, als Erklärung für die beobachteten Wirkungen in Frage kommt. Um dies zu erreichen, muss eine Studie grösstmögliche Objektivität, das heisst eine hohe interne Validität aufweisen. Dazu dürfen die Ergebnisse nicht systematisch in die eine oder andere Richtung verzerrt werden, zum Beispiel durch eine gezielte Behandlungszuweisung, ein ungleiches Mass an Betreuung oder durch Suggestionseffekte. Eine solche systematische Verzerrung bzw. einen solchen systematischen Fehler nennt man auch «Bias» [1]. Wie ein Bias durch die Einführung einer Kontrollgruppe, durch die Randomisierung und die Verblindung der Studienbehandlungen reduziert oder verhindert wird, zeigen die folgenden Abschnitte.

Warum kontrolliert?

«Kontrolliert» bedeutet, dass die Effekte einer Behandlung mit jenen einer anderen, bereits als wirksam eingestuften Behandlung oder jenen nach der Gabe von Plazebo verglichen werden. Bei dem am häufigsten verwendeten Design (Parallelgruppendesign) werden zwei oder mehrere Patientengruppen gebildet, von denen (mindestens) eine als Kontrollgruppe fungiert. Ein solches Vorgehen unterscheidet sich massgeblich von offenen Therapiestudien, kumulativen Fallbeobachtungen oder auch von Vergleichen mit früher erhobenen Daten (sogenannten historischen Vergleichen). Bei den zuletzt genannten Studienformen hat sich gezeigt, dass das Ausmass der erzielbaren Wirkung in der Regel überschätzt wird [2]. Die aktive Einführung einer Kontrollgruppe ist wichtig, um drei mögliche Formen eines Bias zu verhindern:

- Ohne Kontrolle fliesst der natürliche Verlauf einer Erkrankung in die Ergebnisse mit ein [3]. Ist die Spontanheilungsrate hoch, lassen sich ohne Kontrollgruppe kaum Aussagen zur Wirksamkeit einer Behandlung machen,

auch wenn diese anscheinend ausgeprägt ist. Ein Beispiel wäre die Untersuchung eines Antibiotikums bei der unkomplizierten Otitis media des Kindes, die eine hohe Selbstheilungstendenz aufweist. Aber auch im umgekehrten Fall, das heisst bei einer progredient verlaufenden Erkrankung, ist die Kontrolle notwendig. Wäre in der britischen Typ-2-Diabetesstudie («UKPDS») keine Diätkontrollgruppe gebildet worden, dann hätten sich alle untersuchten Antidiabetika als nichtwirksam erwiesen: denn unter medikamentöser Therapie zeigte sich eine fortschreitende Verschlechterung der Stoffwechsellage (die aber geringer ausfiel als unter alleiniger Diätbehandlung). Dass auch historische Kontrollen nicht verlässlich sind, haben in den 70er und 80er Jahren des letzten Jahrhunderts offene Antiarrhythmikastudien bei Patienten nach einem Myokardinfarkt demonstriert. Die im Vergleich zu historischen Daten scheinbar verbesserte Infarktüberlebenschance unter Antiarrhythmikagabe war jedoch nicht Ausdruck der Wirksamkeit der verabreichten Arzneimittel, sondern Folge der über Jahre insgesamt verbesserten Therapie des Myokardinfarktes. Erst die 1989 publizierte, plazebokontrollierte «CAST»-Studie («Cardiac Arrhythmia Suppression Trial») zeigte, dass die Verabreichung von Klasse Antiarrhythmika der Klasse 1C im Gegenteil sogar mit einer zwei- bis dreifachen Mortalitätszunahme verbunden ist.

- «Regression to the mean»: Hierbei handelt es sich um ein statistisches Phänomen, das dann auftritt, wenn Patienten nicht zufällig, sondern auf der Basis extrem erhöhter oder verminderter Werte ausgewählt werden [4]. Das ist häufig der Fall, wenn schwerkranke Patienten für die Untersuchung einer neuen Therapie herangezogen werden. Beispiele dafür sind hohe Blutdruck- oder Cholesterin-Ausgangswerte, stark ausgeprägte allergische oder klimakterische Symptome oder eine sehr niedrige Knochendichte. Mit den im Verlauf der Studie vorgenommenen Wiederholungsmessungen ergeben sich mit einer höheren Wahrscheinlichkeit Werte, die im mittleren Streubereich liegen, das heisst, nach hohen Werten zu Beginn zeigen die Folgemessungen im Mittel niedrigere; umgekehrt treten nach ausgewählt niedrigen Werten anschliessend im Mittel höhere auf. Die anscheinend nachweisbare klinische Verbesserung ist also lediglich Ausdruck des Zufalls – ohne Kontrolle würde das Ausmass der tatsächlich vorhandenen Wirkung überschätzt.
- Plazeboeffekte: Für einige Erkrankungen sind starke Plazeboeffekte beschrieben, zum Beispiel bei Depressionen, Schmerzen oder funktionellen Magen-Darm-Beschwerden. Ohne eine vergleichende Kontrollgruppe würde

sich das ermittelte Wirkungsausmass aus der Summe von Arzneimittel- und Placebo-wirkung zusammensetzen. In den letzten Jahren wird allerdings die Existenz eines Placeboeffektes zunehmend hinterfragt. Einer im Jahr 2001 im «New England Journal of Medicine» publizierten Untersuchung lagen solche Studien zugrunde, in denen als Therapiekontrolle sowohl ein Placebo- als auch ein Nichtbehandlungsarm eingesetzt wurde [5]. Überraschenderweise zeigte sich im Mittel zwischen diesen beiden Gruppen kein signifikanter Unterschied. Es ist daher möglich, dass die Bedeutung von Placebo-wirkungen in den letzten Jahrzehnten überschätzt wurde – unter Umständen waren solche Wirkungen auch eine Folge des «Regression to the mean»-Phänomens. Dafür spricht, dass die Patienten mit dem schlechtesten Ausgangszustand die grössten Placeboeffekte aufweisen.

Mit der kontrollierten Studie, also der Einführung von Vergleichen, sind die obengenannten Formen eines Bias nicht zu eliminieren. Der Bias soll jedoch so gleichmässig als möglich auf alle Behandlungsarme verteilt werden. Das Ausmass der tatsächlichen Wirkung einer medizinischen Massnahme ergibt sich durch die Differenz der Ergebnisse aus aktivem Behandlungs- und Kontrollarm.

Warum randomisiert?

Bei der Randomisierung handelt es sich um eine Methode, um die Patienten zufällig auf die Behandlungs- und Kontrollgruppe(n) zu verteilen. Dadurch soll die bestmögliche Gleichheit der Gruppen bezüglich aller bekannten und unbekannt patientenbezogenen Faktoren erzielt werden, die potentiell das Ergebnis einer Studie beeinflussen können. Bei den meisten Studien wird eine sogenannte vollständige Randomisierung vorgenommen: Jeder Patient hat die gleiche Chance, entweder Behandlung A oder Behandlung B zu erhalten [6]. Von grösster Bedeutung ist es, jene Ärzte, die für die Auswahl der Patienten und deren Aufnahme in die Studie verantwortlich sind, vom Prozess der Randomisierung auszuschliessen. Dazu ist es notwendig, dass die Randomisierung verdeckt erfolgt [7]. Die Zuteilung einer Behandlung wird fallweise erst dann durchgeführt, wenn ein für die Studie geeigneter Patient identifiziert worden ist und dieser seine Einwilligung zur Teilnahme erklärt hat. Worin besteht die Notwendigkeit des Randomisierungsprozesses? Auch die Randomisierung dient dem Ziel, einen möglichen Bias zu verhindern. Denn nimmt der Arzt die Behandlungszuteilung vor oder kennt er diese, werden mit hoher Wahrscheinlichkeit Unterschiede in den

Gruppen bezüglich demographischer oder klinischer Variablen auftreten:

- Der Studienarzt könnte eine Auswahl von Patienten vornehmen, weil sie seiner Meinung nach stärker von einer bestimmten Therapie profitieren. So würde er unter Umständen schwerer erkrankte Patienten der vermeintlich wirksameren Therapie zuführen.
- Es ist auch möglich, dass Patienten sich besser für eine bestimmte Therapie eignen. Zum Beispiel, dass Patienten mit einer angenommen höheren Compliance die Behandlung mit den ausgeprägteren Nebenwirkungen erhalten, um das Ausscheiden von Patienten aus der Studie zu minimieren.

Diese Art der gezielten Behandlungszuteilung nennt man Behandlungsallokationsbias. Er ist deshalb von Bedeutung, weil unzureichend randomisierte Studien – oder Studien mit unzureichender Geheimhaltung der Behandlungszuteilung vor den Studienärzten – grössere Behandlungseffekte vortäuschen [8]. Die Verhinderung eines Behandlungsallokationsbias und die auf Zufall basierende Erzeugung von vergleichbaren Gruppen durch die Randomisierung sind somit wesentliche Voraussetzungen für die Durchführung der «beweisenden» statistischen Auswertung.

Wie sieht die Verhinderung dieses Bias durch Randomisierung nun praktisch aus? Prinzipiell würde Würfeln oder das Werfen einer Münze den Ansprüchen einer Zufallsverteilung der Patienten auf die Behandlungsgruppen gerecht werden. Heute wird die Sequenz der Behandlungszuteilung in der Regel computergestützt durch die Erzeugung von Zufallszahlen (z.B. «1» für Behandlung A und «2» für Behandlung B) festgelegt. Die randomisierte Zuteilung der Behandlungen sowie deren Übertragung in ein Schema zu Verpackung und Einsatz der Medikation («Schachtel Nr. 3 für Patient Nr. 1»), muss durch eine aussenstehende Stelle erfolgen, das heisst abgekoppelt vom direkt in die Studie involvierten Personal. Bei monozentrischen Studien kann die Spitalapotheke diese Aufgabe vollständig wahrnehmen und dem Studienarzt die Medikation patientenweise aushändigen. Bei grossen multizentrischen Studien gibt es häufig eine zentrale Randomisierungsstelle, die dem Arzt (oder der Spitalapotheke) telefonisch oder per Fax mitteilt, welche (bereits vorbereitete) Medikation für einen bestimmten Patienten zu verwenden ist. Somit wird über viele Studienzentren ein balancierter Patienteneinschluss erzielt.

Um einen Behandlungsallokationsbias zu verhindern, muss die Randomisierung für die gesamte Phase des Patienteneinschlusses verdeckt bleiben. Deshalb reicht es nicht aus, dass der Arzt die Behandlungszuteilung anhand einer vorbereiteten Liste vornimmt, da er bereits vor Einschluss der nächsten Patienten wüsste, welche Behandlungen diesen zugeteilt sind. Er

könnte daraufhin Patienten zielgerichtet für die vorgesehene Behandlung auswählen. Die alternierende Behandlungszuteilung («ABABAB ...») verbietet sich ebenfalls, da der Studienarzt nach der «Entblindung» auch nur eines Patienten über den gesamten Zuteilungsschlüssel informiert wäre. Auch die Aushändigung von einzelnen, mit konkreten Anweisungen versehenen Briefumschlägen (z.B. «Patient Nr. 1 erhält Behandlung A») ist anfällig für eine Korrumpierung des Settings: dann nämlich, wenn der Arzt die Umschläge nicht seriell, sondern bereits alle vor Beginn der Studie öffnet und damit Kenntnis der Zuteilungsordnung erhält [9].

Zwei weitere Begriffe sind im Zusammenhang mit der Randomisierung von Bedeutung: die «Blockbildung» und die «Stratifizierung» [10]. Da die Patienten durch die Randomisierung nicht alternierend, sondern zufällig den Gruppen zugeordnet werden, sind diese zu bestimmten Zeitpunkten ungleich verteilt, indem zum Beispiel nach Einschluss von 15 Patienten 5 Patienten auf die Gruppe A und 10 Patienten auf die Gruppe B entfallen. Durch die Bildung von Blöcken, beispielsweise zu jeweils 4 Patienten, wird gewährleistet, dass nach Abschluss eines Blockes jeweils gleichviele Patienten in den Behandlungsgruppen sind (zum Beispiel «ABAB», «AABB», «BAAB» oder «BABA» – aufgrund des Vertauschens der Behandlungsreihenfolge spricht man auch von «permutierter Blockbildung»). Allerdings wird die rein zufällige Verteilung der Patienten auf die Behandlungsgruppen damit partiell aufgehoben (auf «ABA» folgt beispielsweise automatisch «B», auf «AA» folgt «BB»). Ein möglicher Patientenselektionsbias lässt nur dann vermeiden, wenn die Studie doppelblind und dem Studienarzt der Blockmechanismus nicht bekannt ist. Es können auch verschiedene Blocklängen zum Einsatz kommen, die randomisiert über den Studienverlauf verteilt werden. Damit soll die «Vorhersehbarkeit» der nächsten Behandlung – die bei partieller Entblindung und bekannter Blocklänge potentiell gegeben ist – unmöglich gemacht werden.

Eine Stratifizierung wird vorgenommen, um bestimmte Eigenschaften oder Variablen, die das Ergebnis verzerren können oder für die Auswertung der Studie bedeutsam sind, in gleicher Weise auf die Gruppen zu verteilen. So kann zum Beispiel nach Studienzentren (bei multizentrischen Studien), nach Geschlecht oder indikationsspezifisch (prä- vs. postmenopausal, vor- vs. unbehandelt oder Tumorgrösse $\leq x$ cm vs. $> x$ cm) stratifiziert werden. Um dies zu erreichen, werden in der Regel separate Randomisierungsböcke gebildet und die Patienten je nach Vorhandensein oder Abwesenheit einer Eigenschaft randomisiert einem Stratum zugeordnet. Eine Stratifizierung verstärkt den Randomisierungseffekt. Die pro Studienzentrum vorgenommene Randomisierung in Strata zum Rezeptorstatus

(positiv/negativ) und Alter (\leq oder > 50 Jahre) – nur um ein Beispiel zu geben – bedeutet pro Zentrum die randomisierte Zuordnung in vier Untergruppen. Kommt ein weiteres Stratum dazu, sind es pro Zentrum bereits acht. Daran sieht man, dass eine Stratifizierung nur bei ausreichend hohen Patientenzahlen sinnvoll ist.

Über das Beschriebene hinaus existieren alternative Randomisierungsverfahren, die ebenfalls zur Vermeidung von Imbalancen in den Gruppen zur Anwendung kommen, vor allem bei Studien mit relativ wenig Patienten. Bei diesen Verfahren ist die Wahrscheinlichkeit für die Zuordnung eines Patienten zu den verschiedenen Behandlungsgruppen allerdings nicht mehr gleich (wie bei der vollständigen Randomisierung), sondern hängt jeweils von den aktuell bestehenden Unterschieden zwischen den Gruppen bezüglich der Patientenanzahl oder der Verteilung vorab definierter prognostischer Faktoren ab. Wir haben die Einschränkung der zufallsbasierten Behandlungszuordnung bereits bei der Blockbildung kennengelernt. Ein weiteres, häufiger in der Fachliteratur genanntes Verfahren ist die sogenannte «Minimierung» [11]. Dieses und andere alternative, mit Einschränkungen verbundene Randomisierungsverfahren («restricted randomisation») sollen jedoch Gegenstand eines anderen Fortbeitrages sein.

Warum doppelblind?

Die beiden in der Literatur auftauchenden Begriffe «Verblindung» und «Maskierung» werden synonym gebraucht (im Rahmen einer ophthalmologischen Studie sollte man vorzugsweise von Maskierung sprechen). In diesem Beitrag wird nur der Begriff der Verblindung verwendet. Mit Verblindung soll über den gesamten Studienverlauf verdeckt werden, welcher Patient welche Behandlung erhält. Bei einer einfachblinden Studie kennt nur der Patient die ihm zugewiesene Behandlung nicht, bei einer doppelblinden Studie sind Patienten und Studienärzte, meistens auch die Auswerter klinischer Befunde sowie die datenauswertenden Statistiker, «verblindet» (semantisch korrekt müsste es daher eigentlich «tripel-» bzw. «quadrupelblind» heissen). Damit wird bereits deutlich, dass die Verblindung bis zum Abschluss der Datenauswertung aufrechterhalten ist und nicht bereits mit der letzten Behandlung abgeschlossen sein darf. Verblindung ist nicht mit verdeckter Randomisierung zu verwechseln; letztere dient der Vermeidung eines Behandlungsallokationsbias und erfüllt ihren Zweck, sobald alle Patienten in die Studie eingeschlossen sind. Die Verblindung dagegen hält diese verdeckte Zuteilung bis zum Studienende aufrecht. Es versteht sich von selbst, dass die verdeckte Randomisierung die Grundvoraussetzung für die Verblindung der Studie ist.

Eine Verblindung wird am häufigsten im Rahmen pharmakologischer Interventionen vorgenommen. Dazu stehen verschiedene technische Möglichkeiten zur Verfügung [12]. Bei einer plazebokontrollierten Studie erhält die Kontrollgruppe ein Plazebopräparat, das sich bezüglich Aussehen, Farbe, Gewicht und Geschmack nicht vom Präparat mit dem aktiven Wirkstoff unterscheidet. Dem Plazebo kann zum Beispiel auch eine pharmakologisch aktive Substanz zugegeben werden, welche Nebenwirkungen hervorruft, die mit jenen des aktiven Präparates vergleichbar sind. Werden zwei aktiv wirksame Arzneimittel miteinander verglichen, können diese in identisch aussehende Kapseln gefüllt werden. Dieses Verfahren ist jedoch technisch nicht immer möglich, beispielsweise, wenn die Originalpräparate für das Verkapseln zu gross sind. Ausserdem ist diese Methode anfällig für eine Entblindung, etwa wenn (un)beabsichtigt die Kapseln geöffnet werden. Alternativ steht die sogenannte «Double Dummy»-Technik zur Verfügung, bei der für jedes Arzneimittel gleichaussehende Plazebos produziert werden; der Patient erhält dann zur Gewährleistung der Verblindung zu jedem Einnahmezeitpunkt gleichzeitig eine Arzneiform mit aktiver Substanz und das Plazebo zum Vergleichspräparat. Auch zur Untersuchung von verschiedenen Applikationsformen wird die «Double Dummy»-Technik eingesetzt; in bestimmten Situationen, etwa bei einem Vergleich zwischen einer oralen Arzneiform und einer intramuskulären Injektion, ist dieses Verfahren allerdings ethisch fragwürdig (Plazeboinjektionen). Stattdessen könnte eine unabhängige Drittperson die Behandlung vornehmen, ohne dass das direkt in die Patientenführung und Datenauswertung involvierte Personal von der Behandlungszuteilung Kenntnis erhält. Dies ist ein organisatorisch aufwendigeres Verfahren, das zudem ebenfalls mit der Gefahr einer versehentlichen Entblindung einhergeht.

Die Verblindung beginnt jedoch nicht erst mit der Aufbereitung und Verabreichung einzelner Arzneiformen, sondern bereits mit der Primärverpackung der Studienmedikation. Alle Schachteln, Flaschen oder anderen Behälter dürfen sich nicht voneinander unterscheiden und nur über einen Code oder eine Identifikationsnummer differenzierbar sein. Es ist auf jeden Fall zu vermeiden, dass die Medikation gemäss der Gruppenaufteilung in zwei Pakete für Behandlung A und Behandlung B verpackt ist. Die Entblindung nur eines Patienten würde dann automatisch zu einer Entblindung der gesamten Studie führen. Deshalb wird in der Regel nur patientenweise verpackt und ein Nummerierungssystem zur Identifikation verwendet, das dem Arzt die Zuordnung der Medikation zu einer Studiengruppe unmöglich macht.

Warum wird mit der Verblindung ein so hoher Aufwand betrieben, der eine Studie noch kom-

plexer gestaltet, als diese oftmals ohnehin schon ist? Wiederum sollen verschiedene Formen eines Bias verhindert werden. Die in diesem Zusammenhang relevanten systematischen Fehler sind umso bedeutender, je mehr subjektive Faktoren in das Studienergebnis einfließen können [13]. Weiss ein Patient, ob er eine wirksame oder unwirksame Behandlung erhält, wird er in Abhängigkeit davon wahrscheinlich unterschiedlich reagieren. Subjektive Faktoren betreffen jedoch nicht nur den Patienten allein, sondern auch Studienärzte und Datenauswerter. Aus diesem Grund muss in der Regel eine doppelte Verblindung erfolgen. Denn analog zur Randomisierung zeigen nicht oder unvollständig verblindete Studien im Mittel stärkere Behandlungseffekte als doppelblinde Studien [8]. Welche Formen eines Bias spielen hier eine Rolle und sollen durch die doppelte Verblindung vermieden werden?

- **Behandlungsbias.** Ein Behandlungsbias tritt auf, wenn eine besondere Betreuung oder zusätzliche medizinische Massnahmen vorzugsweise einer Behandlungsgruppe zugute kommen [14]. Patienten in der Plazebogruppe erhalten unter Umständen bewusst eine supportive Therapie. Oder die Patienten mit der vom Studienarzt favorisierten Studienbehandlung erfahren eine bessere medizinische Versorgung. Allein die sorgfältigere oder intensivere Führung der Patienten in einer Gruppe kann Suggestionseffekte auslösen, wenn sich der Glaube des Arztes an die Wirksamkeit der Therapie unmittelbar auf den Patienten überträgt.
- **«Attrition Bias» [14]:** Die Entscheidung des Studienarztes, bei einem Patienten aufgrund der Verschlechterung des klinischen Zustandes oder wegen Nebenwirkungen die Studie zu beenden, kann im Wissen der zugrundeliegenden Therapie unterschiedlich ausfallen. Auch bei Patienten, die von sich aus eine Studie abbrechen möchten oder nicht mehr zu einer oder mehreren Studienvisiten erschienen sind, wird man in Kenntnis der Behandlung möglicherweise unterschiedliche Anstrengungen unternehmen, um sie von einer weiteren Teilnahme an der Studie zu überzeugen. Dieser Bias ist deshalb von Bedeutung, weil sich die Patientenzahlen pro Gruppe am Ende der Studie erheblich unterscheiden können. Ausserdem sind Drop-outs nicht in gleicher Weise repräsentativ für die Patientenpopulation, weil sie oftmals andere Charakteristika als die in der Studie verbleibenden Patienten aufweisen. Sie verzerren somit gegebenenfalls die Ergebnisse.
- **«Assessment Bias»:** Ist der Studienarzt von der (Nicht)Wirksamkeit einer bestimmten Therapie überzeugt, könnte dies seine Interviewtechnik bei der Ermittlung oder Bewertung von subjektiven Symptomen des Patienten

ten beeinflussen, zum Beispiel bei der Beschreibung der Schmerzintensität oder der Dokumentation von Nebenwirkungen. Aber auch die Charakterisierung des klinischen Zustandes, Scorebildungen und Stadieneinteilungen könnten bei Kenntnis der Behandlung unterschiedlich vorgenommen werden. Selbst Ultraschallbilder oder Röntgenaufnahmen unterliegen möglicherweise einer abweichenden Interpretation. Dies geschieht nicht absichtlich, sondern kann unbewusst vor dem Hintergrund des allgemeinen Wissensstands, eigener früherer Befunde oder der Plausibilität bzw. fehlenden Nachvollziehbarkeit der erhobenen Daten erfolgen [15]. Ein Beispiel soll das untermauern: In einer plazebo-kontrollierten Studie wurde die Wirksamkeit von Cyclophosphamid und eines Plasmaaustausches bei Patienten mit Multipler Sklerose untersucht. Der klinische Zustand der Patienten wurde sowohl unverblindet als auch verblindet beurteilt. Während nach Einschätzung der «unverblindeten» Neurologen eine Verbesserung auftrat, war diese bei verblindeter Bewertung der Befunde jedoch nicht nachweisbar [16]. Weiterhin kann auch die Entscheidung, ob oder wie Studienprotokollverletzer in die Auswertung einbezogen werden, davon abhängen, ob die zugrundeliegende Behandlung bekannt ist oder nicht.

Grenzen der randomisierten, kontrollierten Doppelblindstudie

Nicht immer kann eine randomisierte, kontrollierte Doppelblindstudie durchgeführt werden: Neue Therapien werden bei Patienten mit fortgeschrittenen Tumoren oftmals zunächst ohne Kontrolle eingesetzt; eine Randomisierung ist bei epidemiologischen Fragestellungen nicht möglich, ebenso eine Verblindung bei Vergleichen zwischen pharmakologischen und chirurgischen Interventionen. Die wenigsten Studien beschreiben im Detail, wie eine Randomisierung vorgenommen wurde oder ob diese oder auch die Verblindung tatsächlich erfolgreich waren [17, 18]. Mit der Randomisierung ist ausserdem keine Garantie verbunden, dass die Behandlungsgruppen identisch sind. Studien, in denen randomisiert zwei verschiedene chirurgische

Verfahren durch den gleichen Chirurgen zur Anwendung kommen, stehen in der Kritik, weil dieser die beiden Verfahren oftmals nicht gleich gut beherrscht. Eine Lösung könnte darin bestehen, dass die Patienten randomisiert entweder einem Experten bezüglich Verfahren A oder einem Experten bezüglich Verfahren B zugewiesen werden («Expertise based randomised controlled trial») [19].

Darüber hinaus ist die randomisierte, kontrollierte Studie selbst anfällig für andere Formen des Bias: Diese betreffen die sogenannte externe Validität, das heisst die Repräsentativität der Ergebnisse. Um eine hohe Homogenität der Patientenpopulation zu erreichen, werden manchmal bis zu 90% der potentiell für eine Studie geeigneten Patienten von vornherein ausgeschlossen (Patientenselektionsbias) [20]. Es ist daher zum Teil fraglich, inwieweit sich solche Ergebnisse auf die Gesamtpopulation übertragen lassen. So sind Herzinsuffizienzpatienten in klinischen Studien durchschnittlich jünger und hauptsächlich männlich; weiterhin weisen sie weniger Begleiterkrankungen auf, und sie haben viel häufiger eine isolierte systolische Dysfunktion, weil der Studieneinschluss von der Auswurfraction abhängig gemacht wird. Auch die Konzentration auf spezialisierte Studienzentren, die ein hohes Niveau der Behandlung und des Patientenmonitoring gewährleisten, kann Studienergebnisse verzerren (Zentrumsselektionsbias). Im Vergleich zur «wirklichen Welt» überschätzen manche randomisierten und kontrollierten Studien daher das tatsächlich vorhandene Wirkungsausmass einer Behandlung deutlich [21].

Schlussbemerkung

Im Vergleich zu anderen Studienformen weist die randomisierte, kontrollierte Doppelblindstudie – trotz ihrer Limitationen – den höchsten methodologischen Standard auf. Sie ist mit der grössten Objektivität verbunden, am wenigsten anfällig für einen Bias und gilt daher zu Recht als Goldstandard im Rahmen der Arzneimittelentwicklung und der evidenzbasierten Bewertung medizinischer Massnahmen. Die randomisierte, kontrollierte und doppelblinde Studie ist ein wesentlicher Faktor zur Aufrechterhaltung der Glaubwürdigkeit klinischer Forschung.

Literatur

- 1 Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004;58:635–41.
- 2 Sacks H, Chalmers TC, Smith H. Randomised versus historical controls for clinical trials. *Am J Med* 1982;72:233–40.
- 3 The James Lind Library. Why comparisons are essential [essay]. www.jameslindlibrary.org/essays/comparisons/comparison.html.
- 4 Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ* 2003;326:1083–4.
- 5 Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *NEJM* 2001;344:1594–602.
- 6 Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209.
- 7 Altman DG, Schulz KF. Concealing treatment allocation in randomised trials. *BMJ* 2001;323:446–7.
- 8 Schulz KF, Chalmers I, Hayes R, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.

- 9 Schulz KF. Subverting randomisation in controlled trials. *JAMA* 1995;274:1456–8.
- 10 Altman DG, Bland JM. How to randomise. *BMJ* 1999;319:703–4.
- 11 Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ* 2005;330:843.
- 12 Schulz KF, Grimes DA. Blinding in randomised trials. Hiding who got what. *Lancet* 2002;359:696–700.
- 13 Day SJ, Altman DG. Blinding in clinical trials and other studies. *BMJ* 2000;321:504.
- 14 Jüni P, Altman DG, Egger M. Systematic reviews in health care. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
- 15 Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453–5.
- 16 Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomised, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;44:16–20.
- 17 Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Contr Clin Trials* 2004;25:143–56.
- 18 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
- 19 Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. *BMJ* 2005;330:88–92.
- 20 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials. Exclusions and selective participation. *J Health Serv Res Policy* 1999;4:112–21.
- 21 Kennedy HL. The importance of randomised clinical trials and evidence-based medicine: a clinician's perspective. *Clin Cardiol* 1999;22:6–12.

Korrespondenz:

Dr. med. Peter Kleist
PFC Pharma Focus AG
Chriesbaumstrasse 2
CH-8604 Volketswil
peter.kleist@pfc.ch