

Schlupflöcher in den Stichproben randomisierter Studien: Ausgeschlossene, Verlorene und Abtrünnige

Kenneth F. Schulz und David A. Grimes

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA

aus: *Lancet* 2002;359:781–5 (Originaltitel: „Sample size slippages in randomised trials: exclusions and the lost and wayward“)

Zusammenfassung

Eine korrekte Randomisierung ist nur wenig wert, wenn nicht alle randomisierten Studienteilnehmer in die Primäranalyse eingeschlossen werden können. Dies ist etwa der Fall, wenn Studienteilnehmer die Nachbeobachtung versäumen, in eine andere Stadt ziehen oder anstelle des verordneten Aspirins Aspartam einnehmen. Der Ausschluss von Studienteilnehmern vor der Randomisierung führt zwar nicht zur Verzerrung des Therapievergleichs, kann aber die Übertragbarkeit der Studienergebnisse beeinträchtigen. Die Eignungskriterien für eine Studie sollten eindeutig und spezifisch sein und vor der Randomisierung zur Anwendung kommen. Der Leser sollte abzuschätzen versuchen, ob eines dieser Kriterien vielleicht dazu führt, dass die Studienstichprobe für die interessierende Patientengruppe untypisch oder nicht repräsentativ wird. Die Bewertung von Studienausschlüssen nach Randomisierung ist an sich einfach: Sie sind nicht zulässig. In die Primäranalyse sollten alle in die Studie aufgenommenen Teilnehmer eingeschlossen und als Teil der ihnen ursprünglich zugeteilten

Studiengruppe analysiert werden (sog. Intention-to-treat-Analyse). In der Realität kommen Verluste allerdings häufig vor. Man sollte deshalb ausreichende Ressourcen für die Entwicklung und Implementierung von Maßnahmen bereitstellen, um eine möglichst große Anzahl von Studienteilnehmern „bei der Stange zu halten“. Ferner sollten die Studienleiter – z. B. durch Verwendung eines Studienprofils – klare und explizite Informationen über die weitere Entwicklung aller randomisierten Teilnehmer während der Studie zur Verfügung stellen. Darüber hinaus können auch Sekundäranalysen durchgeführt werden (etwa Per-Protocol- oder As-treated-Analysen). Solche Analysen sollten jedoch stets als sekundäre und nicht-randomisierte Vergleiche ausgewiesen werden. Der fehlerhafte Umgang mit Studienausschlüssen verursacht schwerwiegende methodische Probleme. Dass manche Erklärungen für den falschen Umgang mit Studienausschlüssen auf den Leser intuitiv attraktiv wirken, verschleiert die Schwere der Problematik, und kreatives Missmanagement in dieser Frage kann die Validität einer Studie untergraben.

wurde? Leider erscheinen viele ungeeignete Hergehensweisen an derartige Probleme auf den ersten Blick logisch und vermeintlich attraktiv. Und genau darin liegt ihre Tücke, weil ungeeignete Verfahren zu schwerwiegenden Verzerrungen führen können. Im vorliegenden Artikel wollen wir uns mit der Frage befassen, wie sich der Ausschluss von Teilnehmern von einer Studie vor und nach der Randomisierung auswirkt.

Ausschluss von Teilnehmern vor der Randomisierung

Teilnehmer können vor der Randomisierung von einer Studie ausgeschlossen werden. Der randomisierte Therapievergleich bleibt davon letzten Endes unbeeinträchtigt (gute interne Validität), unabhängig davon, ob Studienärzte für den Ausschluss bestimmter Personen nun gute oder eher fadenscheinige Gründe ins Feld führen können. In diesem Stadium können Studienausschlüsse allerdings die Extrapolation, d. h. die Übertragbarkeit, der Ergebnisse (externe Validität) beeinträchtigen. Für die Mehrzahl der Untersuchungen empfehlen wir deshalb, die Eignungskriterien ganz im Sinne großer und einfach angelegter Studien auf ein Mindestmaß zu beschränken [3, 4]. Stichhaltige Gründe für den Ausschluss

Eine korrekte Randomisierung [1, 2] ist nur wenig wert, wenn man nicht alle Teilnehmer, die den Studiengruppen randomisiert zugeteilt wurden, auch in die Primäranalyse einschließen kann. Ein wesentlicher Gesichtspunkt bei der Bewertung einer randomisierten, kontrollierten Studie betrifft deshalb die von der Studie ausgeschlossenen oder

ausgeschiedenen Teilnehmer, also Studienausschlüsse, Studienabbrüche, Verluste sowie Abweichungen vom Studienprotokoll. Wie soll man nun mit Studienteilnehmern verfahren, die die Teilnahme an der Studie verweigern, Nachbeobachtungstermine versäumen, die Stadt verlassen oder Aspartam nehmen, obwohl ihnen Aspirin verordnet

bestimmter Teilnehmer gibt es aber durchaus. So leiden manche Personen womöglich an einer Krankheit, für die eine bestimmte Intervention kontraindiziert ist, oder es gibt gute Gründe dafür anzunehmen, dass bestimmte Teilnehmer für die Nachbeobachtung (Follow-up) wahrscheinlich verloren gehen. Richtschnur sollte in jedem Fall die Fragestellung der Studie sein [5]. Manchmal werden aber auch so viele Eignungskriterien aufgestellt, dass die Studie auf eine Population abzielt, für die sich kaum einer sonst interessiert; abgesehen davon, dass auch die Rekrutierung zum Problem werden kann. Werden zu viele oder aber die falschen Teilnehmer aus der Studie ausgeschlossen, so kann dies dazu führen, dass ihre Ergebnisse für die interessierende Personengruppe nicht repräsentativ sind, auch wenn die randomisierte, kontrollierte Studie ansonsten akribisch durchgeführt wurde – d. h. die Ergebnisse könnten wahr, dennoch aber potenziell irrelevant sein.

Worauf man beim Ausschluss von Studienteilnehmern vor der Randomisierung achten sollte

Die Eignungskriterien sollten für die Population charakteristisch sein, über die man Schlussfolgerungen ziehen möchte. Wollen Leser die Ergebnisse einer Studie bewerten, sollten sie sich daher vergewissern, dass die Eignungskriterien eindeutig und spezifisch sind. Besonders wichtig ist, dass diese Kriterien vor der Randomisierung Anwendung finden. Ferner sollte der Leser abzuschätzen versuchen, ob eines der Eignungskriterien womöglich dazu geführt hat, dass die Studienstichprobe untypisch, nicht repräsentativ oder für die interessierende Personengruppe irrelevant ist. In der Praxis allerdings sind die Ergebnisse einer Studie nur selten vollkommen belanglos: „Meistens sind die Unterschiede zwischen unseren Patienten und den Studienpatienten mehr quantitativer Art – d. h. sie unterscheiden sich im Alter oder in ihrem sozialen Status, oder sie haben ein unterschiedlich hohes Risiko für das Zielereignis oder sprechen unterschiedlich auf die

Therapie an) denn qualitativer Art – d. h. es ist überhaupt kein Therapieansprechen zu verzeichnen, oder es besteht keinerlei Risiko für das Zielereignis.“ [6] Solche qualitativen Unterschiede im Therapieansprechen sind selten; die externe Validität von Studien ist tendenziell folglich auch eher robust [6].

Ausschluss von Teilnehmern nach der Randomisierung

Werden Teilnehmer nach ihrer Randomisierung von der Studie ausgeschlossen, besteht die Gefahr einer Verzerrung der Therapievergleiche. Durch die Randomisierung selbst werden bei Studienbeginn unverzerrte Vergleichsgruppen gebildet. Jeder im Verlauf der Studie auftretende „Schwund“ in den ursprünglich unverzerrten Gruppen führt unweigerlich zu Verzerrungen (Bias), abgesehen natürlich von einem zufallsbedingtem „Schwund“, der aber eher unwahrscheinlich ist. Methodiker empfehlen deshalb für die Primäranalyse, die Ergebnisse aller den Studiengruppen randomisiert zugeteilten Patienten auszuwerten, und zwar als Teil der Gruppe, der sie eingangs zugeteilt wurden [3, 7]. Man spricht in diesem Fall von einer Intention-to-treat-Analyse. Einfach ausgedrückt bedeutet das: einmal randomisiert, immer wie zugeteilt analysiert.

Zur Vermeidung von Verzerrungen, die mit einem nicht-zufallsbedingtem Verlust von Studienteilnehmern assoziiert sind, beruht die Primäranalyse in einer randomisierten, kontrollierten Studie auf eben diesem Intention-to-treat-Prinzip [8–10]. Darüber hinaus können natürlich auch Sekundäranalysen durchgeführt werden. Vorzugsweise sollten diese aber bereits im Vorfeld geplant worden sein und zum Beispiel nur diejenigen Teilnehmer berücksichtigen, die das Studienprotokoll vollständig eingehalten haben (Analyse gemäß Studienprotokoll, Per-Protocol-Analyse) oder unabhängig von der randomisierten Zuteilung behandelt worden sind („On-treatment“- oder „As-treated“-Analyse). Sekundäranalysen sind ak-

zeptabel, solange sie auch als sekundäre und nicht-randomisierte Vergleiche ausgewiesen sind. Probleme entstehen allerdings, wenn Studienleiter Studienteilnehmer ausschließen, im Endeffekt aber einen sekundären, nicht-randomisierten Vergleich als den primären randomisierten Vergleich ihrer Studie vorlegen. Realiter ist eine solche Analyse eine Kohortenstudie, die in der Maske einer randomisierten, kontrollierten Studie daherkommt. Der Ausschluss von Studienteilnehmern aus der Analyse kann zu irreführenden Schlussfolgerungen führen (s. Kasten 1) [11–14].

Häufig versäumen die Autoren es, ausreichende Informationen zu den ausgeschlossenen Studienteilnehmern zur Verfügung zu stellen [7, 15, 16]. In einem Review von 249 randomisierten, kontrollierten Studien, die 1997 in wichtigen allgemeinmedizinischen Fachzeitschriften publiziert wurden, stellte man fest, dass nur 2% (fünf von 249) der Berichte explizit angaben, dass alle randomisiert zugeteilten Teilnehmer auch gemäß ihrer randomisierten Gruppenzuteilung analysiert worden waren [17]. In etwa der Hälfte der Studienberichte (119 von 249) wurde die Durchführung einer Intention-to-treat-Analyse angegeben; in vielen Fällen fehlten jedoch nähere Angaben, die diese Behauptung hätten stützen können.

Kasten 1: Eine randomisierte, kontrollierte Studie über Sulfinpyrazon versus Placebo zur Rezidivprophylaxe bei Myokardinfarkt

In dieser Studie führten die Studienautoren eine Primäranalyse durch, in der sie die auf kardiale Ursachen zurückgehenden Mortalitätsraten miteinander verglichen anstatt die Raten aller kardialen Todesfälle [11, 12]. Probleme wurden in ihrer Analyse durch unangemessene Studienausschlüsse verursacht, die letztlich vorgenommen wurden, nachdem sich Patienten als ungeeignet erwiesen hatten [13]: Die Studienärzte hatten sieben Patienten als ungeeignet ausgeschlossen, die eine Behandlung erhalten hatten – sechs in der Behandlungsgruppe und einen in der Placebogruppe, was dazu führte, dass mehr verstorbene Patienten aus der Behandlungsgruppe ausgeschlossen wurden als aus der Placebogruppe. Darüber hinaus lassen die Ergebnisse einer eingehenden Überprüfung dieser Studie

durch die US-amerikanische Food und Drugs Administration (FDA) darauf schließen, dass auf der Basis ähnlicher Kriterien weitere Patienten der Placebogruppe als ungeeignet hätten eingestuft werden können, was jedoch nicht geschehen war [13]. Weiterhin wurde der Ausschluss ungeeigneter Patienten nach ihrer Rekrutierung im Studienprotokoll nicht erwähnt; das galt vor allem für Patienten, die verstorben waren. Die Studienautoren hatten auch zwei Todesfälle in der Sulfipyrazon-Gruppe und einen Todesfall in der Placebogruppe wegen mangelhafter Compliance der Teilnehmer als nicht analysierbar ausgeschlossen, obwohl der Ausschluss von Patienten wegen mangelhafter Compliance im Studienprotokoll nicht vorgesehen war. Ferner benutzten die Studienärzte eine 7-Tage-Regel: Als nicht analysierbar wurden danach alle Todesfälle von Patienten gewertet, die nicht mindestens sieben Tage behandelt worden waren oder die mehr als sieben Tage nach Beendigung der Therapie verstorben waren. Dieses Vorgehen zog allerdings keine allzu heftige Kritik der FDA-Prüfkommission nach sich, was im Wesentlichen wohl darauf zurückzuführen war, dass die 7-Tage-Regel zum einen im Studienprotokoll beschrieben worden war und zum anderen nur einen geringfügigen Gesamteffekt auf die Studienergebnisse hatte. Insgesamt jedoch führten diese unangemessenen Studienausschlüsse zu einer Beeinträchtigung der Studienergebnisse [13]. Obwohl den ursprünglichen Angaben der Studienautoren zufolge die kardial bedingte Mortalität bei den Teilnehmern, die das Medikament einnahmen, um 32% ($p = 0,058$) gesenkt worden war, erbrachte eine Reanalyse ein schwächeres Ergebnis. Wurden die Daten der als ungeeignet oder nicht analysierbar gewerteten Personen in den Gruppen analysiert, in die sie ursprünglich randomisiert worden waren, betrug die Mortalitätssenkung nur noch 21% ($p = 0,16$). Bemerkenswert ist, dass nur die p -Werte mitgeteilt wurden. Wir fordern dringend dazu auf, bei der Präsentation von Studienergebnissen auch die Konfidenzintervalle anzugeben [14]. Darüber hinaus lassen die negativen Auswirkungen unzulässiger Studienausschlüsse, wie von der FDA bekräftigt, Zweifel hinsichtlich der Studie aufkommen. Die Gutachterkommission der FDA erklärte, dass Sulfipyrazon nicht als Medikament zur Vermeidung von Todesfällen in den kritischen Monaten nach einem Herzinfarkt bezeichnet und nicht als solches beworben werden darf, weil die Daten bei genauer Untersuchung nicht so überzeugend waren, wie man auf den ersten Blick vermuten konnte.

Außerdem machen die Autoren häufig überhaupt keine Angaben zu Studienausschlüssen [7]. Diese Informationslücke verleitet viele Leser zu der Annahme, dass die Durchführung der betreffenden Studien auf der Grundlage des Intention-to-treat-Prinzips erfolgte und keine Teilnehmer ausgeschlossen wurden. Wir nennen ein solches Szenario „keine offenkundigen Ausschlüsse“. Im Allgemeinen betrachten Leser solche Studien „ohne offenkundige Ausschlüsse“ als weniger verzerrt, auch wenn es in vielen dieser Studien in Wahrheit wahrscheinlich zu Ausschlüssen gekommen ist, die im Studienbericht nur nicht erwähnt wurden. Tatsächlich zeigten Studien ohne offenkundige Ausschlüsse größere methodische Schwächen als solche, in denen zumindest einige Ausschlüsse von Teilnehmern angegeben worden waren [7]. Mit anderen Worten: Manche stärker verzerrten Studien werden mitunter fälschlicherweise als unverzerrt interpretiert, während viele der weniger verzerrten Studien als verzerrt eingeschätzt werden. Diese Folgewidrigkeit nennen wir Ausschlussparadoxon. Solange Autoren über nach der Randomisierung stattgefundenen Studienausschlüsse nicht umfassend Aufschluss geben, sollte sich der Leser dieses beunruhigende Paradoxon stets vor Augen halten.

Worauf man beim Ausschluss von Studienteilnehmern nach der Randomisierung achten sollte

Bevor wir uns mit der Frage befassen, wie mit nach der Randomisierung erfolgten Studienausschlüssen korrekt zu verfahren ist, wollen wir es nicht versäumen, den Leser darauf hinzuweisen, auf welch dünnes Eis wir uns mit einer solchen Diskussion begeben. Die Berichterstattung über Studienausschlüsse ist mangelhaft, und das Ausschlussparadoxon führt den Leser in die Irre. Studienleiter sollten klare, explizite Informationen über die Entwicklung aller randomisierten Teilnehmer im Verlauf ihrer Studie machen. Fehlen solche Informationen, sollte der Leser misstrauisch werden. Entsprechende Leitlinien

sind in spezifizierten Flussdiagrammen des CONSORT-Statements enthalten [18, 19].

Im Idealfall würden Studienleiter nach der Randomisierung natürlich keinen Teilnehmer mehr ausschließen und eine Intention-to-treat-Analyse durchführen. Studienausschlüsse nach Randomisierung sind denkbar einfach zu bewerten: Sie sind schlicht nicht zulässig. Alle in die Studie aufgenommenen Teilnehmer sollten als Teil der Studiengruppe analysiert werden, in die sie ursprünglich randomisiert wurden. Die medizinische Forschung ist normalerweise zwar nicht so simpel, das Prinzip stimmt jedoch. Um den Ausschluss von Studienteilnehmern nach der Randomisierung auf ein Minimum zu begrenzen, sollte man mit der randomisierten Zuteilung praktischerweise bis zum letztmöglichen Augenblick warten. Findet die Randomisierung bei der ersten Identifizierung des Teilnehmers, aber noch vor Einleitung der Therapie statt, dann müssen alle Ausschlüsse aus der Studie, die vor der Behandlung erfolgen, immer noch als Ausschlüsse nach der Randomisierung gewertet werden. Dieses potenzielle Problem könnte man umgehen, indem man die Randomisierung bis unmittelbar vor Therapiebeginn aufschiebt [20]. Geben die Studienautoren solche Ausschlüsse nach der Randomisierung an, sollten diese gründlich unter die Lupe genommen werden, da sie eine Verzerrung der Studienvergleiche bewirken können. Der Ausschluss von Studienteilnehmern nach der Randomisierung kann mehrere Gründe haben, darunter etwa die Feststellung der Nichteignung eines Patienten; nach Randomisierung, aber vor Therapiebeginn eintretende Ereignisse; Abweichungen vom Studienprotokoll sowie Verluste bei der Nachbeobachtung.

Feststellung der Nichteignung von Studienteilnehmern

In manchen Studien wird erst nach der Rekrutierung der Teilnehmer festgestellt, dass manche von ihnen die Eigenkriterien nicht erfüllen. Der Aus-

schluss von Teilnehmern zu diesem Zeitpunkt kann zu schwerwiegenden Verzerrungen der Studienergebnisse führen, da diese Entdeckung wahrscheinlich nicht zufällig erfolgt. So könnte es beispielsweise vorkommen, dass Studienteilnehmer, die auf die Behandlung nur sehr schlecht ansprechen oder bei denen Nebenwirkungen auftreten, mehr Beachtung finden und deshalb womöglich als eher ungeeignet eingestuft werden. Oder aber ein Arzt, der für bestimmte Teilnehmer eine bestimmte Therapie bevorzugt, schließt Personen von der Studie aus, wenn sie in die seiner Meinung nach falsche Gruppe randomisiert wurden. Teilnehmer, die sich erst nach der Randomisierung als ungeeignet erweisen, sollten in der Studie verbleiben. Ausnahmen sind aber zulässig, wenn sich die Festlegung der Eignungskriterien schwierig gestaltet. In solchen Fällen könnte der Prüfarzt zum Zeitpunkt der Randomisierung von jedem Patienten die gleichen Informationen einholen und diese von einem (gegenüber der Behandlungszuteilung verblindeten) Dritten (eine Person oder Gruppe) zentral überprüfen lassen. Diese Person oder Gruppe könnte Patienten, die die Eignungskriterien nicht erfüllen, dann von der Studie ausschließen, vermutlich ohne dadurch eine Verzerrung der Ergebnisse zu verursachen.

Nach Randomisierung, aber vor Therapiebeginn eintretende Ereignisse

Gelegentlich berichten Studienautoren, dass Studienteilnehmer auf der Grundlage von Ergebnissen ausgeschlossen wurden, die eingetreten sind, bevor die Therapie eingeleitet werden konnte oder eine Wirkung hätte zeigen können. So wurden zum Beispiel in einer klinischen Studie, in welcher der spezifische Effekt eines Arzneimittels auf die Mortalität untersucht wurde, die Daten aller Patienten, die nach der Randomisierung, aber vor Therapiebeginn bzw. vor Ablauf einer mindestens siebentägigen Behandlung verstorben waren, als nicht analysierbar ausgeschlossen. Die-

se Datentrennung scheint intuitiv attraktiv zu sein, da dann keiner der Todesfälle auf die Therapie selbst zurückzuführen ist. Dasselbe Argument könnte aber auch dazu benutzt werden, um die Daten aller Patienten aus Placebogruppen, die während der Studiendauer versterben, von der Analyse auszuschließen, da rein theoretisch keiner dieser Todesfälle mit der Behandlung in Zusammenhang stehen kann. Dieses Beispiel veranschaulicht die potenzielle Eigenwilligkeit im Umgang mit nach der Randomisierung, aber vor Therapiebeginn eintretenden Outcomes.

Randomisierung sorgt auf lange Sicht für eine Balancierung der nicht zuzuordnenden Todesfälle. Jegliches „Herumdoktern“ nach der Randomisierung, und sei es auch noch so wissenschaftlich und unvoreingenommen, kann dieses Qualitätsmerkmal nicht verbessern, sondern nur beeinträchtigen. Und was noch wichtiger ist: Gelegentlich dient das Herumspielen mit den Daten auch der nachträglichen Erklärung unangemessener Studienausschlüsse.

Zu solchen Post-hoc-Interpretationen kommt es nämlich, wenn Studienleiter sich erst ihre Ergebnisse ansehen und dann für ihre Hypothesen günstige Regeln entwerfen. Angenommen, eine Studienärztin stellt die Hypothese auf, dass ein zur Behandlung einer bestimmten Erkrankung eingesetztes Medikament die krankheitsbedingte Mortalität senkt. Nach der Analyse der Daten stellt sie jedoch fest, dass 14 Todesfälle in der Behandlungsgruppe und zwei Todesfälle in der Placebogruppe auftraten, bevor die Therapie eingeleitet oder das Medikament mindestens sieben Tage lang eingenommen worden war. Also deutet sie die Todesfälle als nicht behandlungsbezogen und entfernt diese Daten aus der Analyse. Ein solches Vorgehen würde die Ergebnisse stark verzerren, auch wenn die Argumentation der Autorin im Studienbericht uns wahrscheinlich als folgerichtig erscheinen würde.

A priori aufgestellt, führen solche Regeln bei der Implementierung von Studien zu Komplikationen; a posteriori führen sie zu verzerrten und ungültigen Ergebnissen. Bei der Beurteilung von

randomisierten, kontrollierten Studien ist es in der Regel unmöglich festzustellen, wann die Aufstellung dieser Regeln erfolgte. Am liebsten lesen wir in Studienberichten zu randomisierten Studien, dass nach der Randomisierung keine Teilnehmer mehr aus der Studie ausgeschlossen wurden. In die Analyse sollten die Daten aller randomisierten Patienten eingehen. Ob nun geplant oder ungeplant: Nicht analysierbare Ergebnisse aus Effizienzgründen auszuschließen, gehört im Rahmen der Analyse von randomisierten, klinischen Studien nicht zu den allgemein akzeptierten Vorgehensweisen [21].

Protokollabweichungen

Abweichungen von der zugeteilten Behandlung kommen in vielen Studien vor. Manchmal wird vorgeschlagen, die Teilnehmer, die erheblich von der zugewiesenen Behandlung abweichen, aus der endgültigen Analyse auszuschließen oder sie nur bis zum Zeitpunkt der Abweichung zu berücksichtigen. Auch wenn ein solches Vorgehen verlockend zu sein scheint, so weist es doch eine schwerwiegende Schwäche auf: „Die Gruppe, die von der einen Therapie abweicht, und die Gruppe, die von der anderen Therapie abweicht, können sich so sehr voneinander unterscheiden, [...] dass der Therapievergleich bei den übrigen Patienten stark verzerrt wird.“ [3]

Nehmen wir beispielsweise einmal an, man wollte herausfinden, ob sich die mit der Einlage eines Intrauterinpressars (IUD) assoziierte febrile Morbidität durch eine Antibiotikaprophylaxe senken lässt. Zu diesem Zweck werden Teilnehmerinnen für die Behandlung mit Antibiotika bzw. Placebo randomisiert (s. Abb. 1). Leider weichen 25% der Patientinnen in der Antibiotikagruppe vom Studienprotokoll ab und nehmen ihre Antibiotika nicht ein. Tatsächlich erhalten diese vom Protokoll abweichenden Patientinnen dieselbe Therapie – d. h. in diesem Fall also keine Therapie – wie die Teilnehmerinnen unter Placebo. Sollen sie deshalb aus der Analyse ausgeschlossen werden? Oder sollen ihre Daten mit denen der

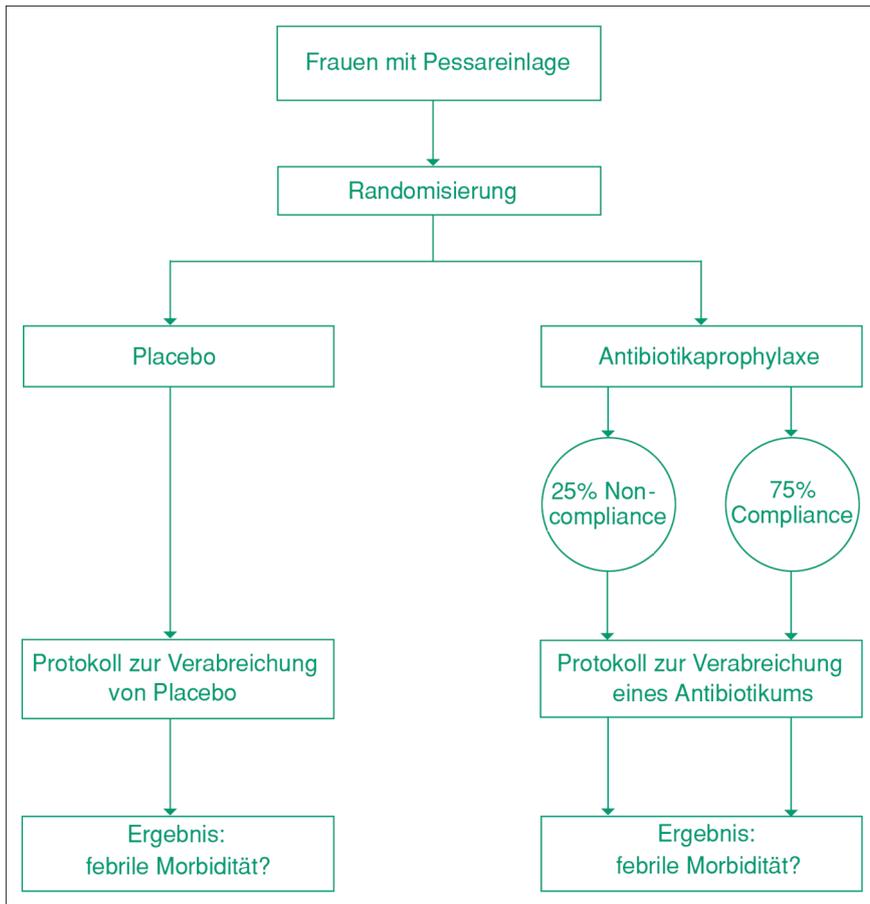


Abb. 1. Schematische Darstellung randomisierter IUD-Patientinnen mit Intrauterinpeppar unter Berücksichtigung ihrer Compliance mit der Behandlung während der Studie (IUD = Intrauterinpeppar).

Placebogruppe zusammengefasst und anschließend mit denen der Patientinnen der Antibiotikagruppe, die sich an das Protokoll gehalten haben, verglichen werden? Und manch ein Studienleiter entscheidet sich dann auch für eine dieser beiden bestechend attraktiven Optionen.

Für die Primäranalyse ist allerdings keine von beiden Möglichkeiten akzeptabel, denn die beiden Studiengruppen sind nicht mehr vergleichbar. Die Teilnehmerinnen, die keine Antibiotika eingenommen haben, waren womöglich gesünder oder haben die Pessareinlage besser toleriert. In jedem Falle waren sie für einen febrilen Infekt wahrscheinlich weniger anfällig. Werden die Daten der vom Protokoll abweichenden Teilnehmerinnen nun aus der Analyse ausgeschlossen, so enthält die Antibiotikagruppe nur noch die anfälligeren Patientinnen. Der Therapievergleich wäre somit verzerrt. Werden ihre Daten zu-

sammen mit denen der Placebogruppe analysiert, dann wären die in der Antibiotikagruppe verbliebenen Teilnehmerinnen nicht nur anfälliger für febrile Morbidität, sondern auch die Placebogruppe wäre von weniger anfälligen Patientinnen „unterwandert“, sodass die Verzerrung des Therapievergleichs sogar noch stärker ausfiele. Die vom Protokoll abweichenden Teilnehmerinnen hätten auch kränker statt gesünder sein können – doch das spielt keine Rolle. Wichtig ist, dass der Therapievergleich systematisch verzerrt wäre.

Alle Protokollabweichungen sollten nachverfolgt und die Daten dieser Studienteilnehmer in der Gruppe analysiert werden, der sie ursprünglich zugeteilt wurden. In unserem Beispiel sollten die vom Protokoll abweichenden Teilnehmerinnen für die Auswertung also in der Antibiotikagruppe verbleiben. Ebenso ratsam ist es, auch die Daten aller vom Protokoll abweichenden Pa-

tientinnen der Placebogruppe in ihrer Gruppe zu analysieren. Ungeachtet dessen, was sich im Laufe der Studie ereignet, sollte man die zur Antibiotikabehandlung randomisierte Gruppe mit der Placebogruppe vergleichen. Dieses Vorgehen beugt nicht nur Verzerrungen vor, sondern gibt eine pragmatische Antwort auf die Frage, die von primärem klinischen Interesse ist – nämlich: Führt die prophylaktische Antibiotikagabe vor einer Pessareinlage zur Vermeidung von febriler Morbidität? Wenn Studienautoren also berichten, dass Teilnehmer mit Protokollabweichungen von der Analyse ausgeschlossen oder von einer Gruppe in die andere verschoben wurden, sollte der resultierende Therapievergleich – in Analogie zu Beobachtungsstudien – als verzerrt eingestuft werden.

Verluste bei der Nachbeobachtung

Verluste in der Nachbeobachtungsphase gehören wohl zu den ärgerlichsten Gründen für den Ausschluss von Studienteilnehmern nach Randomisierung. Es kann immer wieder passieren, dass Teilnehmer umziehen oder die weitere Teilnahme an der Studie verweigern. In der Nachbeobachtung verloren gegangene Teilnehmer könnten aber in der Analyse noch berücksichtigt werden, wenn ihr Behandlungsergebnis aus anderer Quelle in Erfahrung gebracht werden könnte, etwa durch Erhebung der Daten aus einem nationalen Sterberegister. Solche Möglichkeiten ergeben sich allerdings nur selten. Ohne die Ergebnisse dieser in der Nachbeobachtung verlorenen Teilnehmer bleibt den Studienleitern kaum eine andere Wahl, als diese aus der Analyse auszuschließen. Jeder Verlust beeinträchtigt die interne Validität der Studie; aber noch größeren Schaden richten Verlustunterschiede zwischen den Vergleichsgruppen an. Demnach muss Studienleitern daran gelegen sein, die Nachbeobachtungsverluste auf ein Minimum zu beschränken.

In manchen Studien kann die Minimierung solcher Verluste Problemen vorbeugen. Aus diesem Grund sollte man

diesem Punkt ausreichend Beachtung schenken und genügend Ressourcen bereitstellen, um Maßnahmen zur Minimierung solcher Verluste zu entwickeln und zu implementieren [10]. So könnten etwa Patienten, bei denen man davon ausgehen kann, dass sie für die Nachbeobachtung wahrscheinlich verloren gehen, bereits vor der Randomisierung ausgeschlossen werden. Alternativ dazu könnte man auch Kontaktinformationen zum Auffinden verzogener Teilnehmer zu beschaffen versuchen oder spezielle Mitarbeiter für die Nachbeobachtungsphase engagieren, um die nicht zu Nachbeobachtungsterminen erscheinenden Teilnehmer aufzusuchen, oder auch beides tun.

Manche Studienleiter wenden auch innovative Tricks an, um hohe Nachbeobachtungsraten zu gewährleisten, etwa dadurch, dass sie für die Nachbeobachtung mehrere günstig gelegene Kliniken anbieten. Denn nur allzu oft wird von den Teilnehmern erwartet, dass sie eine einzige, oftmals auch noch ungünstig gelegene Klinik zur Nachuntersuchung aufsuchen. Auch die Begrenzung des Datenerhebungsinstrumentariums auf eine überschaubare Größe kommt den Wünschen und Bedürfnissen der Teilnehmer entgegen. Die Follow-up-Raten werden erhöht, wenn man die Teilnehmer nicht überfordert. Womöglich begünstigen solche Instrumente nicht nur höhere Nachbeobachtungsraten, sondern liefern für die wichtigsten Fragestellungen der Studie mitunter auch höherwertige Daten. Verluste komplett auszuschließen ist wahrscheinlich unmöglich, doch viel zu oft berufen sich Autoren in diesem Zusammenhang auf unüberwindliche Schwierigkeiten. Sie könnten sich allerdings stärker darum bemühen, höhere Nachbeobachtungsraten zu erzielen (s. Kasten 2).

Welche Verluste in der Nachbeobachtung können denn nun als akzeptabel gelten? Nur eine Antwort, nämlich 0%, garantiert die Vorteile der Randomisierung. Offenkundig ist dies zuweilen jedoch unrealistisch. Manche Wissenschaftler befürworten daher die einfache Faustregel von „5 und 20“. Nach dieser Regel führen Verluste von weni-

Kasten 2: Maßnahmen zur Maximierung der Teilnahme an der Nachbeobachtung

- Einen Mitarbeiter für die Leitung und Förderung der Nachbeobachtungsphase abstellen
- Mitarbeiter beschäftigen, um Teilnehmer, die nicht zu den Nachbeobachtungsterminen erscheinen, anzurufen oder zu Hause bzw. am Arbeitsplatz aufzusuchen
- Teilnehmer, deren Rückkehr zu den Nachbeobachtungsterminen wahrscheinlich fraglich ist, vor der Randomisierung ausschließen
- Teilnehmer, die wahrscheinlich umziehen, vor der Randomisierung ausschließen
- Kontaktinformationen beschaffen, um Teilnehmer dazu zu bewegen, die Nachbeobachtungstermine wahrzunehmen und das Auffinden dieser Teilnehmer bei Nichterscheinen zu erleichtern, z.B. mittels Postanschrift, Telefonnummer und E-mail von rekrutierten Teilnehmern, engen Freunden oder Angehörigen, die nicht beim Studienteilnehmer wohnen, sowie dem Hausarzt des Studienteilnehmers
- Identifikationsnummer beschaffen (z. B. Krankenversicherungsnummer)
- Für die Nachbeobachtung Orte vereinbaren, die eher den Wünschen der Teilnehmer entgegenkommen als denen der Studienärzte und des Implementierungspersonals, z. B. mehrere in der Nähe der Wohnorte der Teilnehmer gelegene und gut zu erreichende Standorte, nicht nur die zentrale Klinik oder das Zentralkrankenhaus; geringe Wartezeiten anstreben
- Gut organisierte Studienabläufe, um bei Nachbeobachtungsterminen der Studienteilnehmer eine zügige Abwicklung zu gewährleisten
- Datenerhebungsinstrumentarium begrenzen, um die Teilnehmer nicht zu überfordern
- Für eine hervorragende und kostenlose medizinische Versorgung sorgen
- Finanzielle Mittel für Aufwandsentschädigungen und Reisekosten der Teilnehmer bereitstellen

ger als 5% wahrscheinlich nur zu geringfügigen Verzerrungen, Verluste von mehr als 20% stellen hingegen eine potenziell ernste Gefahr für die Validität der Studienergebnisse dar; entsprechend bereiten dazwischen liegende Werte mittelschwere Probleme [22]. Tatsächlich argumentieren die Befürworter dieser Regel – und da stimmen wir mit ihnen überein – aufgrund ihrer Erfahrung mit Sensitivitätsanalysen, der Anwendung des sog. „Worst-case“-Szenarios, dass eine Studie bei Nachbeobachtungsverlusten von mehr als 20% den Gefahren, die der Validität der Studie dadurch drohen, kaum würde erfolgreich standhalten können [6]. Und tatsächlich verweigern manche Fachzeitschriften die Veröffentlichung von Studien, bei denen diese Verluste 20% übersteigen [6].

Auch wenn diese „5 und 20“-Regel Vorteile hat, kann sie in Fällen von seltenen Outcomes zu einer allzu starken Vereinfachung führen [22]. Die für Verluste während der Nachbeobachtung erwarteten Werte hängen von verschiedenen Faktoren wie etwa der unter-

suchten Frage, der Zielereignisrate und der Nachbeobachtungsdauer ab. Würden beispielsweise die Outcomes von Frauen, die in einem Krankenhaus entbunden haben, am ersten Tag nach der Geburt untersucht, bräuchten wir nicht von Verlusten bei der Nachbeobachtung auszugehen. In einer Untersuchung über die Anwendung von Mikrobiziden durch afrikanische Frauen (die normalerweise kein Telefon und manchmal noch nicht einmal eine Straßensanschrift haben) zur Vermeidung einer HIV-1-Übertragung würden wir im Laufe eines einjährigen Nachbeobachtungszeitraums vielleicht mit 5–15%igen Follow-up-Verlusten rechnen müssen, auch wenn wir natürlich auf geringere Ausfälle hoffen würden. Die meisten Studienärzte haben jedenfalls unter solch widrigen Umständen schon viel schlechtere Werte erzielt. Doch ließen sich die Verluste bei der Nachbeobachtung in einer neueren Studie durch verstärkte Anstrengungen auf ca. 1,5% begrenzen [23]. Einer weiteren nützlichen allgemeinen Faustregel zufolge sollten die Verluste bei der Nachbeob-

achtung die Zielereignisrate nicht übersteigen.

Wichtiger vielleicht noch als die absoluten Gesamtverluste bei der Nachbeobachtung sind die Verlusten im Gruppenvergleich. Die Daten sollten hinsichtlich der unterschiedlichen Verlustraten in den jeweiligen Gruppen analysiert werden. Bias könnte dadurch entstehen, dass die Verluste auf Unterschiede bezüglich Unannehmlichkeiten, Toxizität oder Therapiewirksamkeit zurückgeführt werden. In jedem Falle wird empfohlen, die Ergebnisse von Studienteilnehmern, die in der Nachbeobachtung verloren gingen, zu dokumentieren und auszuwerten, und zwar zumindest bis zu dem Zeitpunkt, an dem sie aus der Studie ausgeschieden sind.

Schlussfolgerungen

Man sollte sich darum bemühen, die Zahl der nach Randomisierung von der Studie ausgeschlossenen Teilnehmer auf ein Minimum zu begrenzen und Intention-to-treat-Analysen durchführen. Ferner sollten sich die Studienberichte an den entsprechenden CONSORT-Leitlinien orientieren [18, 19]. Flussdiagramme (Studienprofile) sind besonders hilfreich, um die Entwicklung der Teilnehmer im Studienverlauf verfolgen zu können. Wenn Studienautoren es versäumen, über Studienausschlüsse zu berichten, führt das beim Leser zu Interpretationsschwierigkeiten, etwa zum Ausschlussparadoxon, das den Leser hinsichtlich der Qualität einer Studie in die Irre führen kann. Weiterhin verursacht der fehlerhafte Umgang mit solchen Studienausschlüssen schwerwiegende methodische Probleme. Leider erscheinen die in manchen Studienberichten für solche Schwierigkeiten angebotenen Erklärungen dem Leser auf den ersten Blick recht einleuchtend zu sein, was die Ernsthaftigkeit dieser Problematik allerdings verschleiert. Deshalb ist der Leser aufgefordert, nicht nur einer unzureichenden Studienberichterstattung, sondern auch seiner eigenen Intuition den Kampf anzusagen, um auf diese Weise potenzielle Gefahren für die Validität einer Studie aufzudecken.

Unser Dank gilt Willard Cates und David L. Sackett für ihre hilfreichen Kommentare zu einer früheren Version dieses Artikels. Ein Großteil des Materials entstammt einer mehr als 15-jährigen Lehrtätigkeit im Rahmen des Berlex Foundation Faculty Development Course.

Literatur

- [1] Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002a; 359:515–9.
- [2] Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002b; 359:614–8.
- [3] Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, I: introduction and design. *Br J Cancer* 1976;34: 585–612.
- [4] Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, II: analysis and examples. *Br J Cancer* 1977;35:1–39.
- [5] Sackett DL. On some prerequisites for a successful clinical trial. In: Shapiro SH, Louis TA (eds.) *Clinical trials: issues and approaches*. New York: Marcel Dekker, 65–79 (1983).
- [6] Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill Livingstone (2000)
- [8] Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742–4.
- [8] Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991;10: 1595–1605.
- [9] Lewis JA, Machin D. Intention to treat: who should use ITT? *Br J Cancer* 1993; 68:647–50.
- [10] Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167–89.
- [11] The Anturane Reinfarction Trial Research Group. Sulfapyridazine in the prevention of cardiac death after myocardial infarction. *N Engl J Med* 1978; 298:289–95.

- [12] The Anturane Reinfarction Trial Research Group. Sulfapyridazine in the prevention of sudden death after myocardial infarction. *N Engl J Med* 1980; 302:250–6.
- [13] Temple R, Pledger GW. The FDA's critique of the anturanereinfarction trial. *N Engl J Med* 1980;303:1488–92.
- [14] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet* 2002;359:57–61.
- [15] DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306: 1332–7.
- [16] Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984;5: 328–47.
- [17] Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670–4.
- [18] Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports or parallel-group trials. *Lancet* 2001;357:1191–4.
- [19] Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134:663–94.
- [20] Friedman L, Furberg C, DeMets D. *Fundamentals of clinical trials*. St Louis: Mosby (1996).
- [21] Meier P. Anturane reinfarction trial. *N Engl J Med* 1981;304:730.
- [22] Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. New York: Churchill Livingstone (1997).
- [23] Roddy RE, Zekeng L, Ryan KA, Tamoufé U, Tweedy KG. Effect of nonoxynol-9 gel on urogenital gonorrhoea and chlamydial infection: A randomized controlled trial. *JAMA* 2002;287: 1117–22.

Korrespondenzadresse:

Dr. Kenneth F. Schulz
E-Mail: KSchulz@fhi.org

Anmerkung der Redaktion:

Die Übersetzung dieses Artikels erfolgte durch Frau Karin Beifuss (Stuttgart), die fachliche Bearbeitung übernahm Frau Gerta Rücker (IMBI – Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Freiburg). Beiden sei an dieser Stelle sehr herzlich gedankt.