



ELSEVIER
URBAN & FISCHER

Available online at www.sciencedirect.com



Z.ärztl. Fortbild. Qual.Gesundh.wes. (ZaeFQ) 101 (2007) 419–426

Methodik

Reihe Epidemiologi 6: Generierung von Randomisierungslisten in randomisierten Studien: Zufall, nicht Auswahl

Kenneth F. Schulz, PhD*, David A. Grimes, MD

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

Lancet 2002; 359: 515–19

Die randomisierte kontrollierte Studie ist zwar der Goldstandard der klinischen Forschung, doch ist die Randomisierung nach wie vor der dabei vermutlich am wenigsten verstandene Aspekt. Dabei führen Defizite bei der Randomisierung zu Selektionsbias und Störeffekten (Confounding). Wissenschaftler sollten einen weiten Bogen um alle systematischen, nicht-zufälligen Zuteilungsverfahren machen. Die Studienteilnehmer sollten den Vergleichsgruppen auf der Grundlage von Zufallsverfahren zugeteilt werden. Die elementarste der verschiedenen Methoden zur Generierung einer Randomisierungsliste ist die einfache (uneingeschränkte) Randomisierung, die dem wiederholten fairen Münzwurf vergleichbar ist. Zudem wird die einfache Randomisierung, was die Vermeidung von systematischen Fehlern (Bias) anbelangt, von keinem anderen Verfahren übertroffen, und sei es auch noch so komplex und ausgeklügelt. Studienleiter sollten diese Methode deshalb häufiger anwenden als bisher, und der Leser sollte ungleiche Gruppengrößen erwarten und akzeptieren. Etliche andere komplizierte Verfahren der eingeschränkten Randomisierung begrenzen die Wahrscheinlichkeit unerwünschter Fallzahlunterschiede in den Interven-

tionsgruppen. Das am häufigsten zur eingeschränkten Randomisierung angewendete Verfahren ist die Blockrandomisierung. Bei Anwendung dieses Verfahrens sollten Blockgrößen nach dem Zufallsprinzip variiert und größere Blocklängen gewählt werden; dies gilt vor allem für unverblindete Studien. Weitere eingeschränkte Verfahren wie die Urnenrandomisierung vereinen die vorteilhaften Eigenschaften der einfachen mit denen der eingeschränkten Randomisierung, indem sie die Unvorhersagbarkeit zwar größtenteils bewahren, gleichzeitig aber auch ein gewisses Gleichgewicht herstellen. Die Wirksamkeit der geschichteten Randomisierung hängt davon ab, dass man ein eingeschränktes Randomisierungsverfahren verwendet, um die Gruppenzuteilung in den einzelnen Schichten (Strata) zu balancieren. Die Generierung einer angemessenen Randomisierungsliste kostet nur wenig Zeit und Mühe; der Lohn sind wissenschaftliche Genauigkeit und Glaubwürdigkeit. Untersucher sollten entsprechende Ressourcen für die Generierung adäquat randomisierter Studien bereitstellen und ihre Methoden eindeutig darlegen.

„ ... hat man Zufallszuteilung verwendet, kann, wenn die Studie schließlich in Druck geht, selbst der strengste Kritiker nicht sagen, Unterschiede zwischen den Gruppen seien wahrscheinlich auf Voreingenommenheit oder Dummheit zurückzuführen“ [1].

Bis vor kurzem noch gingen Untersucher bei der Planung ihrer Studien formal kontrollierten Versuchsplänen aus

dem Weg (Kasten 1) [2–5]. Mittlerweile jedoch setzt die randomisierte, kontrollierte Studie in der medizinischen Forschung den höchsten methodologischen Standard (Goldstandard) (Kasten 2) [3,6]. Das außergewöhnliche Potenzial randomisierter, kontrollierter Studien zur Reduktion von Bias hängt davon ab, in welchem Maße Studienleiter die Randomisierung als die wichtigste

Technik zur Vermeidung von Verzerrung auch umsetzen können. Auch wenn die Zufallszuteilung (Randomisierung) der Studienteilnehmer den wichtigsten Aspekt einer kontrollierten Studie darstellt [7], ist sie nach wie vor leider auch der am wenigsten verstandene [8,9].

Im vorliegenden Beitrag beschreiben wir das Grundprinzip, das hinter der

*Zuschriften an: Dr. Kenneth F Schulz
E-Mail: KSchulz@fhi.org (K.F. Schulz).



zufälligen Verteilung und den damit verbundenen Implementierungsmaßnahmen steht. Randomisierung hängt in erster Linie von zwei miteinander zusammenhängenden, aber doch getrennten Prozessen ab, nämlich der Generierung einer nicht vorhersagbaren zufälligen Randomisierungsliste und der Geheimhaltung dieser Liste bis zur endgültigen Zuteilung („allocation concealment“, unvorhersagbare Zuteilung). Hier wollen wir uns darauf konzentrieren, wie man eine solche Randomisierungsliste generieren kann. In einem Folgeartikel geht es um die Geheimhaltung der Zuteilung.

Kasten 1: Geschichte der randomisierten, kontrollierten Studie

Im Laufe des 20. Jahrhunderts hat die kontrollierte Studie als das beste Verfahren zur Bewertung alternativer medizinischer Behandlungs- und Präventionsmaßnahmen zunehmend Anerkennung gefunden. R.A. Fisher [2] entwickelte die Randomisierung als Grundprinzip des experimentellen Forschungsdesigns in den 1920er Jahren und wendete dieses Verfahren vor allem in der agrarwissenschaftlichen Forschung an. Gegen Ende der 1940er Jahre wurde die randomisierte, kontrollierte Studie auch in der Medizin mit Erfolg übernommen, was großenteils dem Einsatz von und der Förderung durch Sir Austin Bradford Hill (siehe Abbildung) während seiner Tätigkeit an der *London School of Hygiene and Tropical Medicine* zu verdanken war [3]. Seine Bemühungen gipfelten in der ersten experimentellen [4] und der ersten publizierten [5] Anwendung von Zufallszahlen bei der Zuteilung von Studienteilnehmern. Schon bald danach hat sich die Randomisierung als wichtigstes Verfahren zur Gewährleistung unverzerrter Vergleichsgruppen etabliert.



Figure reprinted with permission from Elsevier (*The Lancet*, 2002, vol. 359, 515–519).

Worauf man bei der Generierung der Randomisierungsliste achten sollte

Nicht-zufällige Methoden in der Maske von Zufallsverfahren

Paradoxerweise haben viele Forscher fraglos eine alles andere als zufällige Vorstellung von Randomisierung [8–10]. Häufig fassen sie willkürliche Ansätze und alternierende Zuteilungen fälschlicherweise als zufällig auf [11]. Manche Forscher betrachten sogar Vorgehensweisen wie etwa die Zuteilung zu Interventionsgruppen auf der Grundlage von Präinterventionstests, die eindeutig im Widerspruch zur Randomisierung stehen, als quasi-randomisiert [12]. Quasi-randomisiert aber ist so etwas wie quasi-schwanger insofern, als sich beide Begriffe jeder Definition entziehen. Tatsächlich öffnet jede Methode, die einer angemessenen Randomisierung nicht gerecht wird, allen möglichen Störeffekten Tür und Tor. Ohne eine adäquat durchgeführte Randomisierung schleichen sich Selektionsbias und Confounding in Studien ein [7,13]. Manchmal stecken Forscher ihre nicht-randomisierten Methoden – vielleicht

unabsichtlich – in ein randomisiertes Gewand. Sie glauben, dass sie randomisiert haben, beschreiben aber eine Methode, die nicht zufällig ist. Die Zuteilung nach Geburtsdatum, Nummer der Krankenakte oder Vorstellungsdatum oder die alternierende Zuteilung sind keine zufälligen, sondern eher systematische Vorgehensweisen. Dennoch behaupteten in einer von uns durchgeführten Studie [10] die Untersucher in 5% (11 von 206) der Studienberichte, sie hätten ihre Teilnehmer durch solche nicht-zufälligen Methoden randomisiert zugeteilt. Überdies werden nicht-zufallsbedingte Methoden wahrscheinlich sehr viel häufiger benutzt als unsere Ergebnisse vermuten lassen, denn 63% (129 von 206) der Studienberichte enthielten keine näheren Angaben zu der Methode, nach der man die Randomisierungsliste generiert hatte [14].

Aus theoretischen wie auch praktischen Gründen können systematische Verfahren aber nicht zu den Randomisierungsmethoden gerechnet werden. So ist etwa der Wochentag, an dem ein Kind geboren wird, in manchen Populationen nicht allein eine Frage des Zufalls [15]. Außerdem führen systematische Methoden nicht zu einer verdeckten Zuteilung. Definitionsgemäß schließt eine systematische Zuteilung eine adäquate Geheimhaltung im Allgemeinen aus, da diejenigen, die die Teilnehmer für die Studie rekrutieren, die Behandlungszuteilung vorher kennen. Wenn Forscher die Durchführung einer systematischen Zuteilung mitteilen – insbesondere wenn diese als randomisiert maskiert ist –, sollte der Leser die Ergebnisse mit Vorsicht genießen, denn ein solcher Irrtum lässt auf eine mangelnde Kenntnis des Randomisierungsprozesses schließen. Eher vertrauen wir den Ergebnissen einer solchen Studie, wenn die Autoren sie korrekterweise als nicht-randomisiert ausweisen und darlegen, wie Störgrößen kontrolliert wurden. In solchen Fällen sollten Wissenschaftler auch das Ausmaß eines potenziell vorliegenden Selektions- und Informationsbias diskutieren und dem Leser dadurch Gelegenheit geben, die Ergebnisse im Lichte des nicht-randomisierten Charakters der Studie und ihrer systematischen Fehler richtig zu beurteilen.

Methoden zur Generierung von Randomisierungslisten

Zur Vermeidung systematischer Fehler sollten die Teilnehmer an einer Studie den Vergleichsgruppen auf der Grundlage eines zufälligen (randomisierten) Prozesses zugeteilt werden. Dazu wenden Studienleiter verschiedenste Randomisierungsmethoden an [16–20], von denen hier die am weitesten verbreiteten beschrieben werden.

Kasten 2: Vorteile der Randomisierung

Die ordnungsgemäße Implementierung eines Randomisierungsmechanismus bietet mindestens drei wichtige Vorteile:

Randomisierung schaltet systematische Fehler bei der Behandlungszuteilung aus.

Vergleiche zwischen verschiedenen Formen medizinischer Interventionen können irreführend sein, wenn nicht durch entsprechende Vorsichtsmaßnahmen sichergestellt wird, dass eine Studie hinsichtlich der Prognose unverzerrte Vergleichsgruppen umfasst. In kontrollierten Präventions- oder Therapiestudien führt Randomisierung zu unverzerrten Vergleichsgruppen, weil sie Selektionsbias und Störeffekte vermeidet. Folglich sind die Vergleichsgruppen von der – bewussten oder unbewussten – Selektion bestimmter Patienten für eine spezifische Intervention unbeeinflusst. Der Begriff Biasvermeidung beinhaltet sowohl die Eliminierung von Bias bei der Entscheidung über die Aufnahme von Teilnehmern in die Studie als auch bei der Zuteilung der Teilnehmer zu einer Behandlung nach erfolgter Aufnahme. Die Untersucher müssen jeden Teilnehmer sofort nach Feststellung seiner Eignung für die Studie ordentlich registrieren, allerdings ohne Kenntnis der Zuteilung. Die Reduktion von Selektionsbias und Störeffekten unterstreicht die besondere Überlegenheit der Randomisierung. Die Randomisierung hat sich

als das beste Studiendesign für Untersuchungen zu kleinen oder mittleren Therapieeffekten durchgesetzt [6].

Randomisierung erleichtert die Verblindung (Maskierung) der Therapie gegenüber Studienleitern, Teilnehmern und Bewertern und ermöglicht auch eine Anwendung von Placebo [3].

Solche Vorgehensweisen verringern systematische Fehler nach der Randomisierung und würden sich nur schwer und vielleicht überhaupt nicht implementieren lassen, wenn die Teilnehmer den Behandlungen nach einem nicht-zufälligen Schema zugeteilt würden.

Randomisierung gestattet die Anwendung der Wahrscheinlichkeitstheorie, um die Wahrscheinlichkeit auszudrücken, dass etwaige Unterschiede im Outcome von Behandlungsgruppen bloß auf Zufall beruhen.

Einfache (uneingeschränkte) Randomisierung

Die einfache Randomisierung lässt sich als „elementar und doch elegant“ beschreiben (Kasten 3) [21]. Auch wenn es sich dabei um das grundlegendste aller Zuteilungsverfahren handelt, das einem wiederholten fairen Münzwurf entspricht, so wahrt diese Methode doch die vollständige Unvorhersagbarkeit jeder einzelnen Behandlungszuteilung. Die einfache Randomisierung wird, was Unvorhersagbarkeit und Vermeidung systematischer Fehler (Bias) betrifft, von keinem anderen Verfahren zur Generierung der Randomisierungsliste übertroffen, und sei es auch noch so komplex und ausgeklügelt [22].

Die Unvorhersagbarkeit der einfachen Randomisierung kann allerdings auch von Nachteil sein [23]. Bei kleinen Stichprobenumfängen kann eine allein nach dem Zufallsprinzip durchgeführte einfache Randomisierung (1 : 1-Zuteilungsverhältnis) in den Gruppen zu starken Unterschieden im Stichprobenumfang führen. Bei einem Gesamt-

stichprobenumfang von etwa 20 würde es bei ca.10% der durch einfache Randomisierung generierten Zuteilungsfolgen zu einem Ungleichgewicht von 3 : 7 oder noch schlechter kommen [24]. Allerdings nimmt dieses Problem mit wachsendem Stichprobenumfang ab. Die Wahrscheinlichkeitstheorie sorgt dafür, dass sich die Behandlungsgruppen hinsichtlich ihrer Größe auf lange Sicht weniger stark voneinander unterscheiden. Bei einer zweiarmigen Studie mit mehr als 200 Teilnehmern wird das Risiko eines deutlichen Ungleichgewichts der Gruppen nahezu bedeutungslos [24]. Bei Zwischenauswertungen könnte es allerdings bei Stichprobenumfängen unter 200 Teilnehmern zu ungleichen Gruppengrößen kommen.

Münzwurf, Würfeln und das Ziehen von zuvor gemischten Karten stellen angemessene Methoden zur Generierung einfacher vollständiger Randomisierungslisten dar. Alle diese manuellen Losverfahren führen theoretisch zu zufälligen Zuteilungsschemata, erweisen sich in der Praxis häufig aber als nicht-zufällig. Die besten Absichten werden durch falsche Vorstellungen von Randomisierung unterlaufen. Das faire Münzenwerfen z.B. randomisiert die Teilnehmer mit gleich großer Wahrscheinlichkeit zu zwei Interventionsgruppen, kann Untersucher aber dazu verleiten, die Ergebnisse eines Wurfes oder einer Würfelserie zu verändern, wenn z.B. reihenweise nur Kopf und niemals Zahl geworfen wird. Viele Untersucher verstehen die Wahrscheinlichkeitstheorie nicht wirklich und nehmen Zufälligkeit deshalb als nicht-zufällig wahr. Zum Beispiel pflegte der verstorbene Chicagoer Baseball-Reporter Jack Brickhouse zu behaupten, dass ein Schlagmann (*Batsman*), der erfahrungsgemäß bei einem Viertel aller Würfe trifft (ein so genannter „0,250 Hitter“) und nach drei Fehlversuchen das Schlagmal zum vierten Mal betritt, bei diesem vierten Versuch „fällig“ sei, d.h. dass er mit Sicherheit einen Treffer landen würde. Jacks Ankündigung „Jetzt ist er fällig!“ stellt eine nicht-zufällige Interpretation von Zufälligkeit dar. Ähnlich verhält es sich mit Paaren, die schon drei Söhne haben und sich

noch eine Tochter wünschen und häufig denken, dass ihr viertes Kind mit Sicherheit ein Mädchen werden wird, obwohl die Wahrscheinlichkeit, dass sie tatsächlich eine Tochter bekommen, noch immer bei etwa 50% liegt.

Ein Kollege führte in einem Seminar regelmäßig vor, welche falschen Vorstellungen von Randomisierung kursieren. Dazu ließ er die Hälfte der Teilnehmer Zuteilungsschemata mit Hilfe einer angemessenen Randomisierungsmethode erstellen, die andere Hälfte sollte etwas Entsprechendes auf der Grundlage ihrer persönlichen Auffassung von Randomisierung entwickeln. Die Studenten, die ein echtes Zufallsverfahren angewendet hatten, warteten häufig mit langen konsekutiven Serien der einen oder der anderen Behandlung auf. Umgekehrt kam dies bei Studenten, die nach eigenem Ermessen vorgegangen waren, überhaupt nicht vor. So traten Jahrgang für Jahrgang die unter den Studenten verbreiteten verzerrten Vorstellungen von Randomisierung zu Tage.

Überdies sind manuelle Losverfahren nicht nur schwieriger zu implementieren, sondern auch nicht überprüfbar. Angesichts der Gefährdung des Zufallsprinzips, der Implementierungsprobleme und der mangelnden Nachverfolgbarkeit empfehlen wir Untersuchern, auf Münzwurf, Würfeln oder das Ziehen von Karten – auch wenn es sich dabei um akzeptable Methoden handelt – besser zu verzichten. Für welche Methode man sich aber auch entscheiden mag, sie muss im Studienbericht in jedem Fall eindeutig dargelegt werden. Wenn keine solche Beschreibung vorliegt, sollte der Leser die Studienergebnisse mit Vorsicht behandeln. Vertrauen dagegen verdienen Randomisierungslisten, bei denen der Autor die Anwendung einer Zufallszahlentabelle oder eines Zufallszahlengenerators (Computer) erwähnt, da diese beiden Optionen unvorhersagbare, zuverlässige, einfache und reproduzierbare Verfahren darstellen, die zudem noch Nachverfolgbarkeit gewährleisten.

Kasten 3: Einfache Randomisierung

Eine nahezu unbegrenzte Anzahl von Methoden bietet sich an, um auf der Grundlage einer Zufallszahlentabelle eine einfache Randomisierungsfolge zu erzeugen [21]. Für eine gleichmäßige Zuteilung zu zwei Gruppen kann man beispielsweise die Richtung, in der man diese Tabelle lesen will, vorher festlegen: aufwärts, abwärts, links, rechts oder diagonal. Anschließend bestimmt man einen beliebigen Ausgangspunkt, zum Beispiel erste Reihe, siebte Zahl:

56 99 20 20 52 49 **05 78 58 50 62 86**

52 11 88

31 60 26 13 69 74 80 71 48 73 72 18

60 58 20

55 59 06 67 02

Um eine gleichmäßige Zuteilung zu erreichen, könnte man den Interventionen A und B beispielsweise die ungeraden bzw. die geraden Zahlen zuordnen. Somit repräsentiert eine Reihe von Zufallszahlen 05, 78, 58, 50, 62, 86, 52, 11, 88, 31 usw. die Zuteilung zu Intervention A, B, B, B, B, B, B, A, B, A usw. Alternativ könnten die Zahlen 00–49 Therapie A und 50–99 Therapie B entsprechen, oder die Zahlen von 00 bis 09 der Gruppe A und 10–19 der Gruppe B, wobei alle Zahlen über 19 ignoriert werden. Tauglich ist jede dieser unzähligen Optionen, vorausgesetzt die Zuteilungswahrscheinlichkeiten und die Untersucher halten sich an das zuvor festgelegte Schema.

Eingeschränkte Randomisierung

Eingeschränkte Randomisierungsverfahren kontrollieren die Wahrscheinlichkeit, mit der man eine Zuteilungsfolge mit unerwünschten unausgewogenen Stichprobenumfängen in den Interventionsgruppen erhält [20]. Anders ausgedrückt: Wenn Forscher gleich große Behandlungsgruppen anstreben, sollten sie sich für die eingeschränkte Randomisierung entscheiden.

Blockbildung

Ziel einer ausgewogenen (eingeschränkten) Randomisierung sind unverzerrte Vergleichsgruppen, aber auch Vergleichsgruppen, die über die gesamte Studiendauer annähernd gleich groß sind [23]. Diese Eigenschaft erweist sich als hilfreich, wenn die Durchführung von Zwischenauswertungen vorgesehen ist. Die Anwendung einer einfachen Randomisierung könnte bei frühen Zwischenauswertungen gelegentlich sehr ungleiche Stichprobenumfänge ergeben – ein Problem, das sich durch Blockbildung verhindern lässt.

Um das Ziel einer ausgeglichenen Randomisierung zu erreichen, greift man am häufigsten auf die Methode der permutierten randomisierten Blöcke (Blockbildung) zurück. Bei einer Blocklänge von 6 werden beispielsweise von jeweils sechs nacheinander rekrutierten Teilnehmern drei Teilnehmer in die eine und drei in die andere Behandlungsgruppe randomisiert. Das Zuteilungsverhältnis kann aber auch ungleich sein, so werden etwa bei einer Blocklänge von 6 mit einem Zuteilungsverhältnis von 2 : 1 pro Block vier Teilnehmer auf die eine und zwei Teilnehmer auf die andere Behandlungsgruppe verteilt. Diese Methode kann problemlos auch auf mehr als zwei Therapien ausgeweitet werden.

Bei Blockbildung kann über die gesamte Studiendauer eine feste Blocklänge gewählt oder die Blocklänge nach dem Zufallsprinzip variiert werden. Wenn man sich in einer nicht-doppelblinden Studie für Blockrandomisierung entscheidet, sollte man die Blocklänge auch wirklich zufallsvariieren, um so die Chance zu verringern, dass die für die Rekrutierung der Teilnehmer verantwortlichen Mitarbeiter den Zuteilungsplan durchschauen [17]. Bei einer fest gewählten Blocklänge – dies gilt vor allem für kleine Blöcke von sechs oder weniger Teilnehmern – könnte die Blockgröße in einer nicht doppelblind angelegten Studie entschlüsselt werden. Wenn die Behandlungszuordnung nach erfolgter Zuteilung bekannt wird, kann die Folge anhand des Musters der zurückliegenden Zuteilungen auf-

geschlüsselt werden. Spätere Zuteilungen ließen sich somit präzise vorhersagen, wodurch es zu einem Selektionsbias kommen könnte, egal wie wirksam die Geheimhaltung der Behandlungszuteilung auch sein mag. Größere Blocklängen, beispielsweise 10 oder 20, anstelle von kleineren Blöcken – vier oder sechs – und die Zufallsvariation der Blocklängen tragen zur Wahrung der Unvorhersagbarkeit bei [17].

Untersucher, die randomisierte, kontrollierte Studien durchführen, greifen häufig auf das Verfahren der Blockbildung zurück. Skepsis sollte beim Leser allerdings aufkommen, wenn die Blockbildung lediglich erwähnt wird. Studienberichte sollten explizite Informationen enthalten zur etwaigen Verwendung von Blockbildung, zum Zuteilungsverhältnis (meist eins zu eins), zum Zufallsauswahlverfahren (z.B. Zufallszahlentabelle oder Computerbasierter Zufallszahlengenerator) sowie zur Blocklänge bzw. den Blocklängen (bei Zufallsvariation).

Zufallszuteilungsregel

Die einfachste Form der Restriktion ist die Zufallszuteilungsregel. Bei einem gegebenen Gesamtstichprobenumfang sind gleiche Gruppengrößen aber nur am Ende der Studie gewährleistet. Im Normalfall ermitteln Untersucher einen Gesamtstichprobenumfang und wählen dann aus dieser Stichprobe nach dem Zufallsprinzip eine Untergruppe aus, die sie der Gruppe A zuteilen; der Rest wird Gruppe B zugeordnet. Für eine Gesamtstudiengröße von 200 beispielsweise werden 100 Kugeln für die Gruppe A und 100 Kugeln für die Gruppe B in einen Hut gelegt; anschließend werden die Kugeln nacheinander ohne Ersetzung zufällig gezogen; ein solches Vorgehen entspricht der Zufallszuteilungsregel. Die Sequenz würde durch 100 zufällige Zuteilungen zur Gruppe A und 100 zufällige Zuteilungen zur Gruppe B entstehen. Bei dieser Methode ergibt sich ein großer permutierter Block für die gesamte Studie; das bedeutet, dass nur am Ende der Studie ein Gleichgewicht bestünde, nicht aber auch im Verlauf der Studie.

Die Zufallszuteilungsregel bewahrt viele positive Eigenschaften der einfachen vollständigen Randomisierung (vor allem, was die statistische Auswertung betrifft), führt aber mit größerer Wahrscheinlichkeit zu einer zufallsbedingten Unausgewogenheit in den Kovariablen (zufälliges Confounding). Übrigens wird dieser Unterschied bei größeren Stichprobenumfängen belanglos [24]. Zudem leidet im Vergleich zur einfachen vollständigen Randomisierung die Unvorhersagbarkeit. Besonders in einer nicht-doppelblind angelegten Studie entsteht hier durch das Erraten der Zuteilungen (vor allem gegen Ende der Studie) Spielraum für das Auftreten von Selektionsbias, was aber natürlich nicht auf die Ebene der permutierten Blockrandomisierung mit kleineren Blockgrößen zutrifft [24,25].

Manchmal wenden Untersucher als Zufallszuteilungsregel die Methode des Kartenmischens („Shuffling“) an; dazu müssen sie den Stichprobenumfang ermitteln, speziell vorbereitete Karten dem Zuteilungsverhältnis entsprechend auf die Behandlungsgruppen aufteilen, die Karten in Umschläge stecken und die Umschläge mischen, um so eine Art von zufälliger Zuteilung ohne Ersetzung zu gewährleisten [23]. Wahrscheinlich benutzen viele Untersucher diese Methode, bezeichnen sie aber nur selten als „eingeschränktes Mischen“ oder als Zufallszuteilungsregel. Stattdessen geben sie an, dass sie Umschläge benutzt oder ein Shuffling durchgeführt haben. Tatsächlich werden die Generierung der Randomisierungsliste und die Geheimhaltung bei der „restricted shuffled“-Methode zusammengefasst und miteinander verschmolzen. Das Mischen (*shuffling*) bestimmt über die Zuteilungsfolge, und das ist nicht optimal. Was jedoch am wichtigsten ist: Bei der Benutzung von Umschlägen steht und fällt die Angemessenheit der „Shuffling“-Methode mit einer adäquaten Geheimhaltung der Behandlungszuteilung [7,8].

Biased-coin- und Urnenrandomisierung

Mit der Biased-coin-Randomisierung erreicht man nahezu die gleichen Ziele

wie mit der Blockbildung, ohne dabei jedoch eine strikte Gleichheit der Gruppengrößen zu erzwingen [16,26]. Dadurch bleibt die für die einfache Randomisierung typische Unvorhersagbarkeit zum großen Teil erhalten. Biased-coin-Verfahren ändern die Zuteilungswahrscheinlichkeit im Studienverlauf, um eventuell auftretende Ungleichgewichte zu korrigieren (Kasten 4). Bei den adaptiven Biased-coin-Verfahren, von denen die Urnenrandomisierung am besten untersucht wurde, wird die Zuteilungswahrscheinlichkeit je nach Größe des Ungleichgewichts geändert. Biased-coin-Verfahren, darunter auch das Urnenmodell, finden in Studienberichten nur selten Erwähnung. Sie sollten eigentlich häufiger angewendet werden. Einen Computer einzusetzen ist einfacher und zuverlässiger als tatsächlich Kugeln aus einer Urne zu ziehen, genauso wie es bei der einfachen Randomisierung einfacher und zuverlässiger ist, auf den Einsatz eines Computers zu vertrauen als tatsächlich eine Münze zu werfen. In unverbundenen Studien, in denen Unvorhersagbarkeit sehr wichtig ist und die Notwendigkeit der Balance die einfache Randomisierung von vornherein ausschließt, ist das Urnenmodell besonders zweckmäßig. Die Unvorhersagbarkeit des Urnenverfahrens übertrifft die der permutierten Blockrandomisierung, und zwar unabhängig davon, ob man sich für Blockverfahren mit fester oder zufällig variierender Blocklänge entscheidet [26]. Wenn der Leser auf eine Biased-coin- oder Urnenrandomisierung stößt, darf er diese Methoden ruhig als angemessene Verfahren zur Generierung einer zufälligen Verteilung betrachten.

Kasten 4: Biased-coin- und Urnenrandomisierung

Biased-coin-Verfahren verändern die Zuteilungswahrscheinlichkeit im Laufe einer Studie, um möglicherweise auftretende Unausgewogenheiten in den Gruppengrößen zu korrigieren. Beispielsweise können Untersucher die einfache Randomisierung mit

gleicher Zuteilungswahrscheinlichkeit – 0,50 : 0,50 in einer zweiarmigen Studie – anwenden, solange die Differenz in der Anzahl der den Behandlungsgruppen zugeteilten Teilnehmer einen zuvor festgelegten Grenzwert nicht überschreitet. Wenn der Unterschied den Grenzwert erreicht, dann erhöht sich die Wahrscheinlichkeit, dass der nächste Teilnehmer der Gruppe mit den wenigsten Teilnehmern zugeteilt wird (z.B. 0,60 : 0,40). Bei richtiger Implementierung kann ein Biased-coin-Verfahren ein Gleichgewicht herstellen, während gleichzeitig die mit der einfachen Randomisierung einhergehende Unvorhersagbarkeit größtenteils erhalten bleibt [16].

Adaptive Biased-coin-Verfahren, von denen das Urnenmodell am besten untersucht wurde, verändern die Zuteilungswahrscheinlichkeit je nach der Größe des Ungleichgewichts [26]. Das Urnenmodell wird als UD (α, β) abgekürzt, wobei α die ursprünglich vorhandene Anzahl an blauen und grünen Kugeln angibt und β die Anzahl der Kugeln, die zusätzlich in die Urne gelegt werden und von entgegengesetzter Farbe sind als die jeweils gezogene Kugel (α und β = jede beliebige rationale nicht-negative Zahl). Im Falle von UD (2, 1) enthält eine Urne also beispielsweise zwei blaue und zwei grüne Kugeln, beginnend mit einer Wahrscheinlichkeit von 0,50 : 0,50 ($\alpha = 2$). Die Kugeln werden nach dem Zufallsprinzip gezogen und stehen für die jeweiligen Behandlungszuteilungen: blau für Therapie A und grün für Therapie B. In die Urne wird nun eine zusätzliche Kugel von der der gezogenen Kugel entgegengesetzten Farbe ($\beta = 1$) gelegt. Wird zuerst eine blaue Kugel gezogen, dann wären nach der ersten Zuteilung zwei blaue Kugeln und drei grüne Kugeln in der Urne. Für die nächste Zuteilung bedeutet dies ein Verhältnis von 0,40 : 0,60. Wird als zweites wieder eine blaue Kugel gezogen, dann wären nach der zweiten Zuteilung zwei blaue und vier grüne Kugeln in der Urne; das Verhältnis für die nächste Zuteilung betrüge somit 0,33 : 0,67.

Pro Zuteilung wird eine neue Kugel gezogen. Die Zuteilungswahrscheinlichkeiten variieren in Abhängigkeit von den vorangegangenen Zuteilungen.

Replacement-Randomisierung

Bei der Replacement-Randomisierung (Ersetzungsstrategie) wird ein einfaches Randomisierungsverfahren für die Behandlungszuteilung so oft wiederholt, bis die gewünschte Balance erreicht ist. Dabei muss der Untersucher für die Ersetzung objektive Kriterien aufstellen. Bei einer Studie mit 300 Teilnehmern könnte man beispielsweise festlegen, dass ein einfaches Randomisierungsschema ersetzt wird, wenn der Unterschied in der Gruppengröße mehr als 20 Probanden beträgt. Wäre der Unterschied zwischen den durch das erste Schema generierten Gruppen größer als 20, dann würde man ein völlig neues einfaches Randomisierungsschema erstellen, das den ersten Versuch ersetzt, und es nach den objektiven Kriterien auf Unausgewogenheit prüfen. Damit würde man so lange fortfahren, bis ein einfaches Randomisierungsschema gefunden wäre, das die Kriterien erfüllt. Auch wenn einem die Replacement-Randomisierung etwas willkürlich erscheinen mag, stellt sie ein adäquates Verfahren dar, solange es vor Studienbeginn implementiert wird. Zudem ist es leicht umzusetzen und gewährleistet neben einer vernünftigen Balance auch die Unvorhersagbarkeit. Als wichtigste Einschränkung der Replacement-Randomisierung bleibt allerdings festzuhalten, dass sie die Balance zur Durchführung von Zwischenauswertungen nicht über die gesamte Studiendauer sicherstellen kann. Auch wenn dieses Verfahren nur selten angewendet wird, ist es doch als früheste Form der eingeschränkten Randomisierung entstanden [20,27].

Geschichtete Randomisierung

Bezüglich der Ausgangscharakteristika der Behandlungsgruppen kann die

Randomisierung ein zufälliges Ungleichgewicht herbeiführen [28]. Manchmal vermeiden Untersucher solche Ungleichverteilungen durch eine vor der Randomisierung stattfindende Stratifizierung nach wichtigen prognostischen Faktoren wie etwa Alter oder Krankheitsschweregrad. In solchen Fällen müssen Untersucher die Einschränkungsmethode angeben (meist Blockbildung). Um die Vorteile einer solchen Stratifizierung ausnutzen zu können, müssen sie eine Form der eingeschränkten Randomisierung durchführen, damit für geschichtete Subgruppen, die durch potenziell relevante prognostische Faktoren definiert sind, getrennte Randomisierungspläne erzeugt werden können. Eine Stratifizierung ohne Einschränkung bewirkt nichts, ist also nichts weiter als eine „Placebo-Stratifizierung“.

Das Verfahren der Stratifizierung in Studien ist methodisch valide und nützlich, doch schränken theoretische und pragmatische Aspekte seine Anwendung auf die Studienplanung ein. Die mit einer solchen Stratifizierung verbundene zusätzliche Komplexität bringt in großen Studien wenig zusätzlichen Gewinn, da durch die Randomisierung ohnehin für ausgeglichene Gruppen gesorgt ist. Wenn aber doch eine Ungleichverteilung entsteht, dann kann der Untersucher statistisch nach diesen (vorzugsweise im Voraus geplanten) prognostischen Variablen adjustieren [28,29]. Das größte Problem bereitet die Tatsache, dass die zusätzliche Komplexität der Stratifizierung die Mitarbeiter womöglich davon abhält, an einer Studie teilzunehmen bzw. die Probanden während des hektischen Klinikbetriebs in die Studie aufzunehmen, was beides die Rekrutierung beeinträchtigt. Für große Studien bietet die Stratifizierung nur unbedeutende Vorteile, ist aber mit relevanten praktischen Nachteilen verbunden. Eine wichtige Ausnahme jedoch stellen multizentrische Studien dar, in denen die Stratifizierung nach Zentrum gewisse Vorteile verspricht, ohne für die Studienleiter in den einzelnen Zentren die Komplexität zu erhöhen. Eine weitere potenzielle Ausnahme ergibt sich ebenfalls im Zusammenhang mit großen

Multicenterstudien, in denen eine zentrale Randomisierung durchgeführt wird. Die zentrale Randomisierung schränkt die praktischen Nachteile der Stratifizierung ein, und in Studienzentren mit kleineren Fallzahlen lassen sich auf diese Weise sogar möglicherweise gewisse Vorteile erzielen.

Eine Stratifizierung kann in kleinen Studien von Nutzen sein, indem sie schwerwiegende Ungleichheiten hinsichtlich der prognostischen Faktoren vermeiden hilft. Sie sorgt für eine angemessene Balance (bezüglich der stratifizierten Faktoren) und wahrscheinlich auch für etwas mehr statistische Power und Genauigkeit [17]. Der aus der Stratifizierung resultierende Gewinn fällt jedoch in sich zusammen, wenn die Teilnehmerzahl pro Gruppe die 50 erst einmal überschritten hat [17]. Ferner kann die Stratifizierung indirekt auch negative Effekte hervorrufen, wenn innerhalb kleiner Schichten ein exaktes Gleichgewicht angestrebt wird. Um eine exakte Balance zu erreichen, greifen Untersucher daher oftmals auf kleine, feste Blocklängen zurück, wodurch wiederum aber auch das Prinzip der Unvorhersagbarkeit verletzt wird. Das Verfahren der Minimierung (*minimisation*) vereint „Stratifizierung“ und „eingeschränkte Randomisierung“ [16]. Minimierung kann bewirken, dass kleine Gruppen im Hinblick auf verschiedene Charakteristika sehr ähnlich werden. Streng genommen muss Minimierung als nicht-zufällig angesehen werden [21]; entscheidet man sich aber für ihre Anwendung, sollte vorzugsweise eine Zufallskomponente implementiert werden. Minimierung findet sowohl Befürworter [30] als auch Kritiker [22]. In jedem Falle müssen Untersucher, die sich dieser Methode bedienen, dafür Sorge tragen, dass diejenigen, die die Studie umsetzen, keine Kenntnis von den bevorstehenden Zuteilungen sowie andere Informationen erhalten, die es ihnen ermöglichen, die als nächstes anstehenden Zuteilungen zu erraten [16].

Trennung von Generierung und Implementierung

Häufig vernachlässigen Studienleiter – meist unabsichtlich – ein weiteres wichtiges Element des Designs und der Berichterstattung einer randomisierten, kontrollierten Studie. Bei allen Verfahren dürfen die Personen, die das Zuteilungsschema entwickelt haben, nicht auch an der Eignungsfeststellung der Teilnehmer, der Verabreichung der Behandlung oder an der Beurteilung des klinischen Ergebnisses beteiligt sein. Denn die betreffenden Personen hätten in der Regel Zugang zum Zuteilungsplan und damit auch die Gelegenheit zur Verzerrung der Ergebnisse [8]. Mängel in dieser Studienphase sind womöglich Schwachstellen, durch die systematische Fehler in die Studie eindringen können. Das CONSORT Statement behandelt unter Punkt 10 (Implementierung) genau dieses Thema [28,31]. Forscher sollten also in ihren Studienberichten angeben, wer die Randomisierungsliste generiert hat, wer die Teilnehmer rekrutiert und wer sie zugeteilt hat. Wer die Randomisierungsliste erzeugt hat, sollte nicht für die Rekrutierung und Zuteilung der Teilnehmer verantwortlich sein. Dennoch kann es unter bestimmten Umständen vorkommen, dass ein Untersucher das Zuteilungsschema erzeugen und auch die Rekrutierung und Zuteilung übernehmen muss. In solchen Fällen muss er dafür sorgen, dass der Zuteilungsplan nicht vorhersagbar ist und ihn für alle, insbesondere sich selbst, unzugänglich aufbewahren.

Fazit

In der medizinischen Forschung setzen randomisierte, kontrollierte Studien den höchsten methodologischen Standard (Goldstandard). Das Schlüsselwort hierbei lautet „randomisiert“, und diese Randomisierung muss angemessen durchgeführt werden. Die Generierung einer Randomisierungsliste erfordert nur wenig Zeit und Aufwand, belohnt uns jedoch mit wissenschaftlicher Genauigkeit und Glaubwürdigkeit. Unter-

sucher sollten entsprechende Ressourcen bereitstellen, um eine korrekte Randomisierungsliste generieren und ihre Methoden klar darlegen zu können.

Wir danken Willard Cates und David L. Sackett für ihre hilfreichen Kommentare zu einer früheren Version dieses Artikels. Ein großer Teil des Materials stammt aus 15 Jahren Lehrtätigkeit im Rahmen des Fakultätsentwicklungsprogramms der Berlex Foundation.

Literatur

- [1] Hill AB. The clinical trial. *N Engl J Med* 1952;247:133–9.
- [2] Fisher RA. *The Design of Experiments*. London: Oliver and Boyo; 1935.
- [3] Armitage P. The role of randomization in clinical trials. *Stat Med* 1982;1:345–52.
- [4] Doll R. Controlled trials: the 1948 watershed. *BMJ* 1998;317:1217–20.
- [5] Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;2:769–82.
- [6] Peto R. Why do we need systematic overviews of randomized trials? *Stat Med* 1987;6:233–44.
- [7] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- [8] Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456–8.
- [9] Schulz KF. Unbiased research and the human spirit: the challenges of randomized controlled trials. *CMAJ* 1995;153:783–6.
- [10] Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;272:125–8.
- [11] Grimes DA. Randomized controlled trials: “it ain’t necessarily so”. *Obstet Gynecol* 1991;78:703–4.
- [12] Grimes D, Fraser E, Schulz K. Immunization as therapy for recurrent spontaneous abortion: a review and meta-analysis. *Obstet Gynecol* 1994;83:637–8.
- [13] Altman DG. Randomisation. *BMJ* 1991;302:1481–2.
- [14] Schulz KF, Chalmers I, Altman DG, Grimes DA, Dore CJ. The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. *Online J Curr Clin Trials* 1995; Doc No 197: [81 paragraphs].
- [15] MacFarlane A. Variations in number of births and perinatal mortality by day of

- week in England and Wales. *BMJ* 1978; 2:1670–3.
- [16] Pocock SJ. *Clinical trials: a practical approach*. Chichester: Wiley; 1983.
- [17] Meinert CL. *Clinical trials: design, conduct, and analysis*. New York: Oxford University Press; 1986.
- [18] Friedman L, Furberg C, DeMets D. *Fundamentals of clinical trials*. St Louis: Mosby; 1996.
- [19] Piantadosi S. *Clinical trials: a methodologic perspective*. New York: John Wiley and Sons; 1997.
- [20] Lachin JM. Statistical properties of randomization in clinical trials. *Control Clin Trials* 1988;9:289–311.
- [21] Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
- [22] Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. *Control Clin Trials* 1988;9:365–74.
- [23] Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol* 1998;41:245–56.
- [24] Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials* 1988;9:312–26.
- [25] Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials* 1988;9:327–44.
- [26] Wei LJ, Lachin JM. Properties of the urn randomization in clinical trials. *Control Clin Trials* 1988;9:345–64.
- [27] Cox DR. *Planning of experiments*. New York: Wiley; 1958.
- [28] 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.
- [29] 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.
- [30] Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998;317:362–3.
- [31] Moher D, Schulz KF, Altman DG, Lepage L. The CONSORT statement: revised recommendations for improving the quality of reports or parallel-group randomised trials. *Lancet* 2001;357:1191–4.

Anmerkung der Redaktion:

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

Neue Mitglieder in die Leitlinienkommission der AWMF berufen

Das Präsidium der AWMF hat bei seiner Sitzung im März beschlossen, die Ständige Kommission Leitlinien der AWMF um vier persönliche Mitglieder zu erweitern. Für die Dauer von 3 Jahren wurden auf Grund ihrer Erfahrungen und Verdienste um die Leitlinienentwicklung berufen:

- Monika Lelgemann MSc. (Klin. Epi.)
Institut für Gesundheits- und Medizinrecht (IGMR), FB Rechtswissenschaft, Universität Bremen

- Prof. Dr. med. Berthold Rzany Sc.M.
Division of Evidence Based Medicine (dEBM),
Klinik für Dermatologie der Charité Berlin
- PD Dr. Karsten Schwerdtfeger
Neurochirurgische Klinik, Universitätskliniken
des Saarlandes Homburg/Saar
- PD Dr. Helmut Sitter
Institut für Theoretische Chirurgie, Philipps-
Universität Marburg
Die Ständige Kommission Leitlinien der
AWMF hatte bereits auf ihrer letzten Sitzung

Neue Themen für Nationale VersorgungsLeitlinien beschlossen

Als nächste Themen für Nationale VersorgungsLeitlinien haben die Träger des Programms für NVL am 9. Mai 2007 Demenz und Adipositas/Metabolisch-vaskuläres Syndrom beschlossen. Bei beiden Erkrankungen handelt es sich um Entitäten mit zunehmender Häufigkeit, großer volkswirtschaftlicher Bedeutung und um einen sektorenübergreifenden Behandlungsbedarf.

Die neu priorisierten Nationalen VersorgungsLeitlinien zu Demenz und Adipositas/Metabolisch-vaskuläres Syndrom sollen einen Beitrag zur Verbesserung der sektorenübergreifenden Behandlung leisten. Die Priorisierung dieser Themen geht auf ein offenes Vorschlagsverfahren zurück. Unter der Adresse <http://www.versorgungsleitlinien.de/kontakt/> llvorschlag

besteht die Möglichkeit, Themen für NVL vorzuschlagen. Die Auswahl erfolgt unter Nutzung allgemein akzeptierter, formaler Priorisierungs- und Konsensus-Verfahren. Die Beschlussfassung erfolgt durch die Planungsgruppe des ÄZQ.

Bisher wurden Nationale VersorgungsLeitlinien zu den Themen Asthma, COPD, KHK sowie Augen- und Fußkomplikationen bei Typ-2-Diabetes fertig gestellt. Sie sind neben weiteren Materialien wie PatientenLeitlinien und speziellen Implementierungshilfen unter <http://www.versorgungsleitlinien.de> im Volltext abrufbar und erscheinen darüber hinaus in gedruckter Form im Deutschen Ärzte-Verlag. Derzeit in Bearbeitung befindlich sind NVL zu den Themen Neuropathie, Nephropathie und makrovaskuläre Erkrankun-

Personalia

beschlossen, diese vier Personen auf Grund ihrer Fortbildung und praktischen Erfahrungen in der Entwicklung von Leitlinien für die nächsten drei Jahre zu offiziellen AWMF-Leitlinien-Beratern zu ernennen. Eine Verlängerung ist möglich.

Quelle: <http://www.egms.de/en/journals/awmf/2007-4/awmf000118.shtml>

Magazin

gen bei Typ-2-Diabetes, Herzinsuffizienz, Kreuzschmerz und Depression.

Korrespondenzadresse:

Sylvia Säger
Dr. Public Health, Dipl.-Ing.
Bereichsleitung Medizinische Information
Ärztliches Zentrum für Qualität in der Medizin (ÄZQ)
Gemeinsames Institut von BÄK und KBV
Wegelystraße 3/Herbert-Lewin-Platz,
10623 Berlin
Tel: 030-4005-2520, Fax: 030-4005-2555
E-Mail: Saenger@azq.de
URL: www.azq.de