

Wer evidenzbasiert argumentiert, bezieht sich auf den aktuellen Wissensstand aus Studien, nicht nur auf persönliche Erfahrungen oder auf die Meinung von Experten. Das bloße Zitat einer Studie ist aber noch keine evidenzbasierte Argumentation. Studien bieten keine unumstößlichen Wahrheiten, sondern Ergebnisse statistischer Analysen. Jeder Studientyp hat dabei spezifische Stärken und Schwächen. Diese Artikelreihe konzentriert sich auf klinische Studien, die experimentell Wirksamkeit und Verträglichkeit von Arzneimitteln prüfen. In kurzen Beiträgen möchten wir Sie mit dem nötigen „Werkzeug“ ausstatten, um klinische Studien zu Arzneimitteln kritisch zu lesen und sich Ihre eigene, evidenzbasierte Meinung zu bilden.

Klinische Studien zu Arzneimitteln – Wo ist der Haken?

Randomisierung – der reine Zufall

Angenommen, Sie sind Hausarzt in einer kleinen Stadt an der Ostsee. Jeden Sommer behandeln Sie zahlreiche Urlauber mit Quallenstichen. Bislang haben Sie den leidgeplagten Patienten immer die Salbe „Medusa“ empfohlen – mit gutem Ergebnis, wie Sie fanden. Doch bei einem Treffen unter Kollegen hören Sie, dass die Salbe „Ictus“ viel wirksamer sein soll. Sie trinken nachdenklich Ihr Glas Wein aus und beschließen, Ihre Erfahrungen experimentell zu überprüfen.¹ Für Ihre kontrollierte Studie möchten Sie zwei Behandlungsgruppen bilden, die sich möglichst ähnlich sind. Wie gelingt das am besten?

Möglichkeit 1: Menschen entscheiden

Ein naheliegender Gedanke wäre, dass Sie selbst die Patienten entweder zu einer Behandlung mit Medusa oder mit Ictus zuteilen. Dafür stellen sie zunächst eine Liste an Merkmalen auf, welche den Krankheitsverlauf oder die Wirksamkeit der Therapie beeinflussen (z. B. Schweregrad des Quallenstichs, Alter, bekannte Allergien). Bei der Zuteilung der Patienten achten Sie darauf, dass sich die beiden Behandlungsgruppen hinsichtlich der zuvor festgelegten Merkmale ähneln.

Wo ist der Haken? Egal wie viele objektive Kriterien Sie sich vorab überlegen – es bleibt ein Entscheidungsspielraum, in dem Sie beeinflusst werden durch Ihre Erwartungen und Wünsche. In unserem Beispiel haben Sie bislang gute Erfahrungen mit Medusa gemacht, Sie erwarten also, dass sich Medusa als wirksamer und verträglicher herausstellen wird als Ictus. Nun sitzt Frau Müller vor Ihnen, die gerade eine palliative Chemotherapie beginnen musste. Ist es nicht verständlich, dass Sie Frau Müller zumindest mit dem lästigen Quallenstich so schnell wie möglich helfen wollen? Frau Mayer dagegen, die ständig über ihre neue Hüfte jammert, aber ansonsten keine schweren Erkrankungen hat – warum sollten Sie es ihr nicht zumuten, dass sie einen Versuch mit Ictus unternimmt? Vielleicht setzt sich aber auch Ihr ärztlicher Ehrgeiz durch und sie wollen den besserwisserischen Kollegen beweisen, dass Sie Ihre Patienten all die Jahre richtig behandelt haben. Wäre es dann nicht naheliegend, dass Sie Frau Müller die Salbe Ictus geben und Frau Mayer die Salbe Medusa? Möglicherweise handeln Sie auch unter dem Eindruck einer Fortbildungsreihe, die durch das Unternehmen Medusamper unterstützt wurde. Die Anwendungsbeobachtungen zu Medusa wurden dort äußerst überzeugend präsentiert. Ihre Studie bietet Ihnen jetzt die Chance, dass nächstes Jahr Sie Ihre „Real-world“-Erfahrungen auf dem

Einhart, N.

Dicheva-Radev, S.

Mathes, T.

¹ Bei diesem Gedankenexperiment klammern wir statistische Fragen zur Errechnung der geeigneten Stichprobengröße ebenso aus wie die erforderliche Genehmigung Ihrer Studie durch die zuständige Bundesoberbehörde sowie das notwendige positive Votum der Ethikkommission in Ihrem Bundesland.

Rückblick: Ohne Kontrolle geht nichts:

In kontrollierten Studien gibt es (mindestens) zwei Studienarme: eine Patientengruppe, die das untersuchte Arzneimittel erhält, und (mindestens) eine weitere Patientengruppe, die entweder keine oder eine andere Therapie erhält. Ein kontrolliertes Studiendesign ermöglicht es, zwischen dem natürlichen Krankheitsverlauf und dem Effekt des Arzneimittels zu unterscheiden – allerdings nur dann, wenn die Patientengruppen in allen Merkmalen übereinstimmen, die den Krankheitsverlauf oder die Wirksamkeit des Arzneimittels beeinflussen.

Kongress vorstellen. Das würde Ihre Kollegen beeindrucken – und mit dem großzügigen Vortragshonorar könnten Sie Ihren Mitarbeitern endlich den gewünschten Kaffeevollautomaten finanzieren. Unabhängig davon, ob Sie persönliche oder finanzielle Interessen haben: Ihre Erwartungshaltungen und Ihr Wunsch nach einem bestimmten Studienergebnis führen (unbewusst) dazu, dass Risikopatienten zwischen den beiden Armen nicht gleichverteilt sind.²

² Es handelt sich hierbei um ein Beispiel zu einem fiktiven Arzt und fiktiven Arzneimitteln. Somit sind auch die geschilderten Gedankengänge, Wünsche und Beweggründe fiktiv. Informationen zum Umgang mit Interessenkonflikten in der Medizin finden Sie hier: [Dtsch Arztebl 2011; 108\(6\): A 256–60.](#)

Confounder (Störfaktoren) beeinflussen sowohl den untersuchten Endpunkt als auch die Intervention, die ein Patient erhält. In der Studie „Medusa versus Ictus“ wird das Merkmal „schwere Vorerkrankung“ zum Confounder, weil dieses Merkmal nicht nur das Ansprechen auf die Therapie, sondern auch die Zuteilung der Intervention beeinflusst (Abbildung 2). Hierdurch entstehen Behandlungsgruppen, die sich in ihrem natürlichen Krankheitsverlauf oder ihrem Ansprechen auf die Therapie unterscheiden. Diese Ungleichheit stört den Vergleich zwischen den Behandlungsgruppen: Es ist unklar, inwieweit die Studienergebnisse auf die unterschiedlichen Interventionen oder aber auf die unterschiedlichen Patientenmerkmale zurückzuführen sind (**Selektionsbias**). Confounder werden am besten vermieden, indem der menschliche Entscheidungsspielraum und der Einfluss äußerer Umstände minimiert wird. Die Zuteilung zu den Behandlungsgruppen erfolgt dann randomisiert, also rein zufällig. Die Randomisierung führt (tendenziell) zu einer **Strukturgleichheit** der Behandlungsgruppen nicht nur hinsichtlich bekannter und leicht messbarer Patientencharakteristika, sondern auch hinsichtlich noch unbekannter (z. B. genetische Merkmale) bzw. schwer messbarer Faktoren (z. B. Compliance).

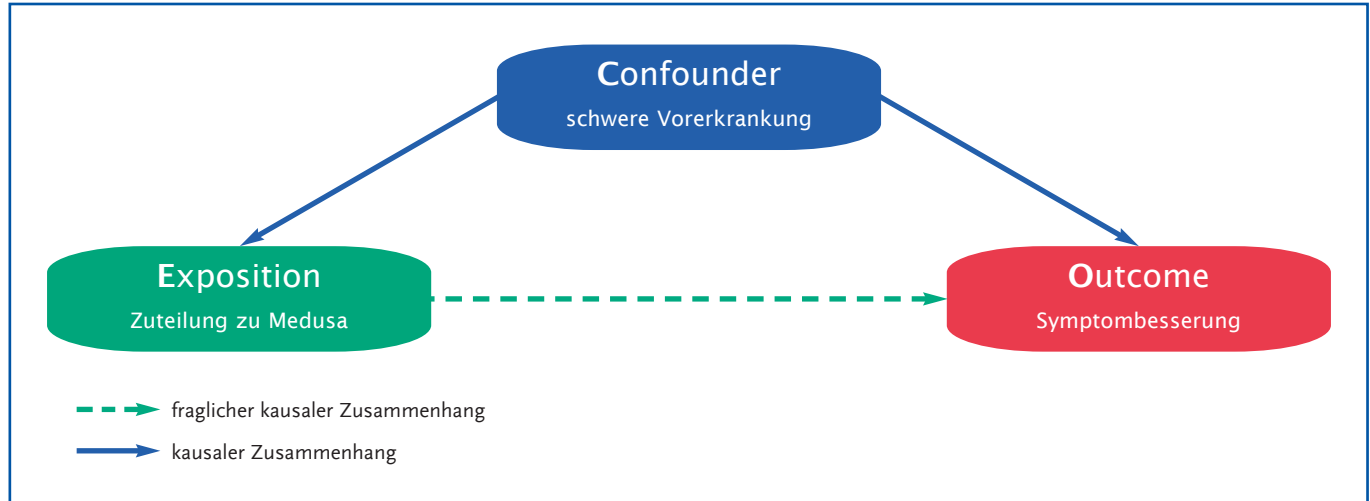


Abbildung 1: Confounding: Eine schwere Vorerkrankung bedingt die Zuteilung zu Medusa und täuscht einen kausalen Zusammenhang zwischen der Behandlung mit Medusa und der Symptombesserung vor. Die schwere Vorerkrankung ist daher ein Confounder.

Möglichkeit 2: Umstände entscheiden

Um sich weder durch Mitleid noch durch Ehrgeiz verführen zu lassen, beschließen Sie, Ihre Patienten in der Vormittagssprechstunde mit Ictus und in der Nachmittagssprechstunde mit Medusa zu behandeln.

Wo ist der Haken? Die Zuteilung anhand eines vorab festgelegten Schemas, das nicht auf Zufall basiert, wird als Quasi-Randomisierung bezeichnet. Dieses Verfahren erhöht im Vergleich zur echten Randomisierung das Risiko für einen Selektionsbias: Erstens besteht die Gefahr, dass die äußeren Umstände selbst zum Confounder werden (siehe Kasten

„Confounder“). So ist es denkbar, dass berufstätige Menschen eher in Ihre Nachmittagsprechstunde kommen, während Menschen, die aufgrund ihres Alters oder chronischer Erkrankungen nicht arbeiten, eher vormittags in Ihre Praxis gehen. Dann verordnen Sie vermehrt jüngeren und gesünderen Menschen Ictus – und die Gleichheit der Behandlungsgruppen ist nicht mehr gewährleistet. Zweitens: Die pfiffige MFA Ihrer Praxis, der gegenüber Sie Ihr Studiendesign angedeutet haben (oder die es durchschaut hat), wird ebenso von Erwartungshaltungen und Wünschen beeinflusst wie Sie. Wenn Frau Müller zufällig ihre Nachbarin ist, liegt es dann nicht nahe, dass Ihre MFA Frau Mayer am Dienstagmorgen in Ihr Sprechzimmer schickt, aber Frau Müller erklärt, dass – leider! – die Sprechstunde so voll sei, dass Sie besser am Nachmittag erneut vorbeischauchen sollte?



© Adobe Stock | #500516906

Abbildung 2: Die Kunst der Randomisierung: Zufallszuteilung, die alle relevanten Unterschiede berücksichtigt.

Möglichkeit 3: Zufälle entscheiden

Sie installieren auf Ihrem Praxiscomputer einen Zufallszahlengenerator. Patienten mit geraden Zahlen erhalten Ictus, Patienten mit ungeraden Zahlen erhalten Medusa.

Wo ist der Haken? Eine zufällig generierte Zahl ist im Unterschied zu äußeren Merkmalen des Patienten (wie sein Erscheinen in der Vormittagsprechstunde) kein Confounder. Allerdings muss garantiert sein, dass Ihre MFA nicht die generierte Zufallszahl und damit die Gruppenzuteilung kennt, da andernfalls eine Manipulation der Patientenreihenfolge möglich wäre, indem sie bestimmte „ungewollte“ Patienten nicht (oder erst zu einem anderen Zeitpunkt) in die Studie einschließt. Allgemeiner gesagt: Die Zuteilung in eine Behandlungsgruppe sollte frühestens offengelegt werden, nachdem der Patient in die Studie

aufgenommen wurde. Eine verdeckte Zuteilung (concealed allocation) ist auch in unverblindeten Studien möglich, denn die offene Gabe der Studienmedikation erfolgt erst *nach* der Zuteilung in die Behandlungsarme. Für die verdeckte Zuteilung werden in der Regel IRT (Interactive Response Technology)-Systeme verwendet.

Interactive Response Technology (IRT)

Bei einer Randomisierung werden die Patienten durch ein Dialogsystem per Zufall auf die Behandlungsarme mit dem Prüfarszneimittel und dem Vergleichsarszneimittel (ggf. Placebo) zugeteilt. Solche Randomisierungssysteme werden Interactive Voice Response Systems (IVRS) oder Interactive Web Response Systems (IWRS) genannt. Das IWRS ist webbasiert. Hier gibt der Prüfer über eine Online-Plattform die für die Studie relevanten Patientendaten ein (z. B. Alter oder Geschlecht). Erfüllt der Patient zudem alle Ein- und Ausschlusskriterien für die Studie, wird ihm über die Datenbank eine Randomisierungsnummer zugeordnet. Damit ist der Patient in der Studie registriert und gleichzeitig randomisiert, d. h. er wurde einem der verschiedenen Behandlungsarme zugeordnet. Früher wurde häufig das IVRS zur Randomisierung von Patienten in klinischen Prüfungen verwendet. Es funktioniert ähnlich wie das IWRS, aber über das Telefon.

Stratifizierung und Blockbildung

Insbesondere bei kleineren Studien können bei der Randomisierung zwei Probleme auftreten: 1. Die Behandlungsgruppen unterscheiden sich zufällig deutlich in ihrer Größe. 2. Die Behandlungsgruppen unterscheiden sich zufällig in relevanten Merkmalen.

Das erste Problem (ungleiche Gruppengröße) kann durch **Blockbildung** verhindert werden. Dafür werden die Patienten zunächst in Gruppen (Blöcke) unterteilt. Die Zuweisung zu dem jeweiligen Behandlungsarm erfolgt anschließend innerhalb eines Blocks. Sie könnten sich z. B. für eine Blockbildung von vier Patienten entscheiden, weil Sie die Erfahrung gemacht haben, dass häufig mindestens vier Patienten mit Quallenstich gleichzeitig im Warteraum sitzen. Von diesen vier Patienten werden dann je zwei der Behandlung mit Medusa und Ictus zugeteilt. Je kleiner die Blocklänge ist, desto größer ist bei unverblindeten Studien die Wahrscheinlichkeit, dass die Therapiezuweisung vorausgesagt werden kann: Wenn bei einer Blockgröße von vier Patienten bereits Frau Müller und Frau Meyer zu Medusa randomisiert wurden, dann wissen Sie mit 100-prozentiger Sicherheit, dass die nächsten beiden Patienten Ictus erhalten werden. Idealerweise wird deshalb die Blockgröße im Studienverlauf zufällig variiert.

Dem zweiten Problem (Strukturungleichheit) kann durch **Stratifizierung** entgegengewirkt werden. Dabei werden die wesentlichen Merkmale identifiziert, die die untersuchte Erkrankung beeinflussen können wie z. B. Alter, Krankheitsstadium oder -schweregrad, Geschlecht und Ethnizität. In Abhängigkeit von der Ausprägung eines Merkmals werden die Patienten in verschiedene Schichten (Strata) aufgeteilt. Für jede Schicht erfolgt eine separate Blockbildung und Randomisierung. Wenn Sie z. B. festgestellt haben, dass Quallenstiche bei Kindern und Jugendlichen schneller verheilen, dann sollten Sie für Ihre Studie nach Alter stratifizieren. Auch könnte die Wirksamkeit der Therapie durch den Zeitpunkt des Behandlungsbeginns beeinflusst sein. Möglicherweise sind am Folgetag des Quallenstichs die Beschwerden bereits am Abklingen. Wenn in einem Studienarm Patienten überwiegen, die ihre Behandlung erst zu einem späteren Zeitpunkt beginnen, so zeigen diese eine

schnellere Symptombesserung als die Vergleichsgruppe – auch dann, wenn sie keine wirksamere Therapie erhalten (zu selbstlimitierenden Symptomen siehe auch Kasten: Das praktische Beispiel).

Tabelle 1: Checkliste Randomisierung

Methodische Frage ^a	Fundort in der Primärpublikation	Beurteilung
Wie wurde das Randomisierungsschema erstellt?	in der Regel im Methodikteil („Methods“) beschrieben, oft in einem eigenen Absatz („Randomization and Blinding“)	Eine IT-gestützt generierte Randomisierungssequenz („computer-generated random sequence“) spricht gegen das Vorhandensein eines Confounders. Dieser Aspekt wird oft nicht explizit beschrieben. Bei Nennung einer IRT bzw. eines „central computerized systems“ ist davon auszugehen, dass das Randomisierungsschema methodisch korrekt erstellt wurde.
Wurde die Zuteilung erst nach Aufnahme des Patienten in die Studie offengelegt?		Die Nutzung von IRT („using interactive response technology“) bzw. eines „central computerized system“ spricht für eine verdeckte Zuteilung. Erfolgt eine Blockbildung, sollte insbesondere bei unverblindeten Studien die Bildung kleiner, fester Blöcke vermieden werden (siehe oben, Blockbildung).
Ähneln sich die Gruppen hinsichtlich relevanter Merkmale, die den Krankheitsverlauf oder das Ansprechen auf die Therapie beeinflussen?	fast immer als Tabelle dargestellt („Baseline Demographics and Clinical Characteristics“)	Die Strukturgleichheit muss klinisch beurteilt werden. Die Angabe von p-Werten bei Patientencharakteristika ist nicht zielführend. ^b

^a Diese Fragen orientieren sich am Risk of Bias (RoB) 2 Tool von Cochrane (1), Das RoB 2 Tool wird in einem späteren Artikel dieser Reihe ausführlich vorgestellt.

^b Bei einer Randomisierung stimmt per definitionem die Nullhypothese („Die Wahrscheinlichkeit für eine Merkmalsausprägung ist in beiden Gruppen gleich“). Zufallsschwankungen bei der Merkmalsausprägung werden bei der Hauptanalyse berücksichtigt.

Das praktische Beispiel: Fezolinetant (Veoza)

Fezolinetant ist zugelassen zur Behandlung von moderaten bis schweren Hitzewallungen, die mit der Menopause assoziiert sind. Der G-BA sah einen Anhaltspunkt für einen geringen Zusatznutzen. Unter anderem begründete der G-BA (2) seine Entscheidung mit den „grundsätzlichen Unsicherheiten“, die sich aus Unterschieden bei den Patientencharakteristika der Studie DAYLIGHT (3) ergeben. Der durchschnittliche Beginn der Amenorrhö lag bei Patientinnen im Fezolinetant-Arm mehr als ein Jahr länger zurück als bei Patientinnen im Placebo-Arm (Gesamtpopulation: 72 Monate vs. 59 Monate (4), für die Nutzenbewertung herangezogene Teilpopulation: 73 Monate vs. 57 Monate (5). Da Hitzewallungen überwiegend etwa 4,5 Jahre nach der letzten Regelblutung andauern (2), ist aus Sicht des G-BA unklar, inwieweit sich mehr Patientinnen im Fezolinetant-Arm zu einem Zeitpunkt in der Studie befunden haben, an dem ihre menopausalen Beschwerden bereits begonnen hatten abzunehmen.

Der Randomisierungsprozess wird in der Primärpublikation (3) nur marginal beschrieben. Im Rahmen der Nutzenbewertung wurden umfangreichere Informationen veröffentlicht (4): Ein unabhängiger unverblindeter Statistiker erstellte ein Block-Randomisierungsschema stratifiziert nach Raucherstatus; die Gruppenzuteilung erfolgte durch IRT, wobei Studienärzte und Patientinnen gegenüber der Zuteilung verblindet blieben. Aus methodischer Sicht ist von einer zufälligen Zuteilung in der Studie DAYLIGHT auszugehen. Insbesondere bei Behandlungsgruppen mit weniger als 1000 Teilnehmern können zufallsbedingt einzelne Merkmale ungleich zwischen den Gruppen verteilt sein. Eine klinisch relevante Strukturungleichheit reduziert die Aussagesicherheit der Studie. Um dem entgegenzuwirken, wäre eine Stratifizierung nach dem klinisch wichtigen Merkmal „Zeit nach Beginn der Amenorrhö“ sinnvoll gewesen.

Literatur

- 1 Cochrane Deutschland, Institut für Medizinische Biometrie und Statistik, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin. Manual zur Bewertung des Biasrisikos in Interventionsstudien. 2. Auflage, 2021. Verfügbar unter: https://freidok.uni-freiburg.de/files/194900/1-z1FUxpKBJdVB3G/Manual_Bewertung_V2.0.pdf.
- 2 Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V): Fezolinetant (Vasomotorische Symptome (VMS), mit der Menopause assoziiert). 1.8.2024. Verfügbar unter: https://www.g-ba.de/downloads/40-268-10704/2024-08-01_AM-RL-XII_Fezolinetant_D-1035_TrG.pdf.
- 3 Schaudig K, Wang X, Bouchard C, Hirschberg AL, Cano A, Shapiro C M M et al. Efficacy and safety of fezolinetant for moderate-severe vasomotor symptoms associated with menopause in individuals unsuitable for hormone therapy: phase 3b randomised controlled trial. BMJ 2024; 387:e079525. doi: 10.1136/bmj-2024-079525.

Fazit

Die randomisierte kontrollierte Studie ist der Goldstandard der klinischen Forschung, aber Defizite bei der Randomisierung können zu Selektionsbias führen. Dadurch kann die Validität der Studienergebnisse – ihre wissenschaftliche Genauigkeit und Glaubwürdigkeit – kompromittiert werden. In klinischen Studien zu Arzneimitteln ist es daher essenziell, die Studienteilnehmer auf der Grundlage von Zufallsverfahren (Randomisierung) den Behandlungsgruppen zuzuteilen. Das genaue Verfahren der Randomisierung sollte in Publikationen immer explizit beschrieben werden. Insbesondere sollte nachvollziehbar sein, auf welche Weise die Gruppenteilung vor Patienten und Studienärzten bis zum Studienbeginn geheim gehalten wurde. Studien ohne Randomisierung oder mit methodisch schlechter Randomisierung liefern verzerrte Ergebnisse und sollten bei Therapieentscheidungen nicht berücksichtigt werden.

Interessenkonflikte

Die Autorinnen und der Autor geben an, keine Interessenkonflikte zu haben.

4 Astellas Pharma GmbH. Dossier zur Nutzenbewertung gemäß § 35a SGB V) Fezolinetant (VEOZA™). Modul 4 A: Behandlung von Frauen mit moderaten bis schweren vasomotorischen Symptomen (VMS), die mit der Menopause assoziiert sind. Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Stand: 25.01.2024. Verfügbar unter: https://www.g-ba.de/downloads/92-975-7459/2024_01_25_Modul4A_Fezolinetant.pdf.

5 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Fezolinetant (vasomotorische Symptome bei Menopause) – Addendum zum Projekt A24-15. Projekt: A24-69; Version 1.0; Stand: 12.7.2024. Verfügbar unter: https://www.g-ba.de/downloads/92-975-7672/2024-08-01_Addendum-IQWiG_Fezolinetant_D-1035.pdf.

Dr. med. Natascha Einhart, Berlin
natascha.einhart@baek.de

Dr. P.H. Stanislava Dicheva-Radev, Berlin

Prof. Dr. Tim Mathes, Göttingen