

Ein- und Ausschlusskriterien bei Therapiestudien: Für wen gelten die Ergebnisse einer Studie?

Prof. em. Dr. med. Johann Steurer

Für Ärztinnen und Ärzte stellt sich nach der Lektüre einer Therapiestudie die Frage: **Auf welche Patientinnen und Patienten sind die Ergebnisse dieser Studie übertragbar?** Eine Antwort auf diese Frage ist relevant, um individuelle Patienten über die Wirksamkeit aber auch die Nebenwirkungen informieren zu können. Inwieweit sich die Ergebnisse auf Personen, die nicht an der Studie teilgenommen haben, übertragen lassen wird mit der **externen Validität** einer Studie beschrieben¹.

Ein einfaches Beispiel: Ergebnisse einer Studie in der z.B. die Wirksamkeit (Verhinderung eines Hirnschlages) eines blutdrucksenkenden Medikaments in die ausschließlich unter 30-Jährige eingeschlossen wurden, lassen sich kaum auf 90-Jährige übertragen. Das ist eine Studie mit sehr eingeschränkter externer Validität, da sich die Ergebnisse nur auf eine begrenzte Patientenpopulation – Hypertonie ist bei unter 30-Jährigen eher selten – übertragen lassen.

Häufig wird im Zusammenhang mit der externen Validität auch der Begriff „Generalisierbarkeit“ verwendet. Dieser Begriff ist jedoch irreführend. Generalisierbarkeit impliziert, dass sich aus einem Ergebnis eine allgemein gültige Regel ableiten lässt. Dies ist in der Medizin jedoch nur selten möglich, da Studienergebnisse meist nicht für alle Patienten mit einer bestimmten Erkrankung gelten. In der praktischen Medizin stehen nicht generalisierte Aussagen über Patientengruppen im Vordergrund, sondern die Frage nach dem individuellen Krankheitsverlauf eines einzelnen Patienten – mit oder ohne die in einer Studie untersuchte Therapie. Präziser ist daher der Begriff der „Extrapolierbarkeit“ der Ergebnisse. Gemeint ist damit die Übertragbarkeit von Studienergebnissen auf individuelle Patienten, die nicht an der Studie teilgenommen haben. Die externe

¹ Die «interne Validität» einer Studie bezieht sich auf die Korrektheit der Ergebnisse. D.h. wurde die Studie «lege artis» durchgeführt um Verzerrungen (Bias) oder falsche Ergebnisse zu verhindern. z.B. wurde die Randomisierungsliste geheim gehalten, wie wurde mit «drop-outs» umgegangen, ...

Validität – also die Übertragbarkeit der Studienergebnisse – wird im Wesentlichen durch die *Ein- und Ausschlusskriterien* einer Studie bestimmt.

Ein- und Ausschlusskriterien

Ein wichtiger Schritt bei der Planung von Therapiestudien (meist randomisierte Studien) ist die Festlegung der Ein- und Ausschlusskriterien und damit wie 'eng' oder 'weit' diese Kriterien gefasst werden sollen. Dies hat Auswirkungen auf die Übertragbarkeit der Ergebnisse in der klinischen Praxis.

Beispiele für Ein- und Ausschlusskriterien, die präzise und verständlich zu beschreiben sind, sind *demographische Charakteristika* – Alter, Geschlecht, Ethnie – *studienpezifische Charakteristika* – Krankheit, Stadium der Krankheit, Dauer der Krankheit, frühere Behandlung der Krankheit, usw. – *und weitere Kriterien* – andere Krankheiten, Nikotinabusus, usw..

Beispiele zur Erläuterung: Aspirin und Remdesivir

Vor rund 20 Jahren wurde Aspirin zur Prävention kardiovaskulärer Erkrankungen empfohlen. Grundlage für die Empfehlung war vor allem das 1989 publizierte Ergebnis der Physicians' Health Study (1). Aspirin (damals in einer Dosis von 325 mg/d) senkt bei Ärzten im Vergleich zu Placebo das Herzinfarktisiko signifikant. In späteren Studien, meist mit einer Tagesdosis von 100 mg (wegen der Blutungskomplikationen bei der höheren Dosis), liess sich dieser Nutzen für die Primärprävention nicht bestätigen. Gleichzeitig rückten die Nebenwirkungen, insbesondere gastrointestinale und zerebrale Blutungen, stärker in den Fokus. Da bei langfristiger Einnahme das Blutungsrisiko auch bei niedriger Dosierung relevant ist, wird Aspirin heute nicht mehr generell zur Primärprävention empfohlen.

Zwei randomisierte, placebokontrollierte Studien² veranschaulichen, welchen Einfluss die Einschlusskriterien auf die Studienergebnisse haben können. Beide Studien schlossen Personen ohne bekannte kardiovaskuläre Erkrankung ein und untersuchten die Wirkung von Aspirin in der Primärprävention.

² Es gibt weit mehr als zwei Studien dazu. Die Ergebnisse der anderen Studien ändern aber nichts an den Schlussfolgerungen der beiden zitierten Studien.

Studie 1 (2): Menschen älter als 55 Jahre (Männer) oder 60 Jahre (Frauen) mit einem moderaten (eher tiefen) kardiovaskulären Risiko. Intervention: Aspirin 100 mg/d oder Placebo. Studiendauer: im Mittel 5 Jahre. Outcome (ein sogenannter composit Outcome): kardiovaskulärer Tod oder Myokardinfarkt oder instabile Angina pectoris oder Schlaganfall oder transient ischämische Attacke. Outcome eingetreten: in der Aspirin Gruppe bei 4.3% und in der Placebo Gruppe bei 4.5% ($p = 0.60$). Die Anzahl gastrointestinaler Blutungen war in der Aspirin Gruppe hingegen mehr als doppelt so hoch als in der Placebo Gruppe.

*Studie 2 (3): Menschen älter als 40 Jahre mit einem **Diabetes mellitus** aber ohne bekannte kardiovaskuläre Erkrankungen. Intervention und Outcome wie in Studie 1. Dauer der Studie: im Mittel 7.4 Jahre. Outcome eingetreten: 8.5 % in der Aspirin Gruppe und 9.6 % in der Placebo Gruppe ($p = 0.01$). Gravierende Blutung bei 4.1 % in der Aspirin Gruppe und 3.2 % in der Placebo Gruppe. Der positive Effekt – Reduktion der kardiovaskulären Ereignisse – wird durch das Risiko gravierender Blutungen aufgehoben.*

Ein weiteres Beispiel stammt aus der Zeit der Covid-Epidemie.

Bei einem Teil der Patienten mit COVID-19 verlief die Krankheit ohne gravierende Beschwerden und Hospitalisationen, ein Teil musste vor allem wegen Atemproblemen hospitalisiert werden, ein Teil starb und ein Teil litt oder leidet bis heute an den Folgen der Infektion (Long COVID).

Ein Medikament zur Behandlung von Patienten mit COVID-19 ist Remdesivir. Remdesivir ist ein Virostatikum, das ursprünglich für die Behandlung der Viruserkrankung Ebola entwickelt wurde. Dieses Virostatikum hemmt die virale RNA-Polymerase und damit die Vermehrung des Virus. Dazu zwei Studien mit unterschiedlichen Ergebnissen.

In der einen Studie (4) wurden Patienten mit COVID-19 eingeschlossen, die nicht hospitalisiert waren und die Symptome nicht länger als eine Woche andauerten – also mit einer «leichteren» Infektion. Remdesivir reduzierte, verglichen zu Placebo, die Hospitalisations- und Sterberate (Hospitalisation

oder Tod: 0.7 % in der Remdesivir Gruppe und 5.3 % in der Placebo Gruppe; Hazard Ratio 0.13; 95%-Konfidenzintervall 0.03-0.59). Remdesivir zeigte einen signifikanten klinischen Benefit.

Und in der anderen (5) Studie wurden Patienten eingeschlossen die wegen COVID-19, verminderter Sauerstoffsättigung und radiologischen Zeichen einer Pneumonie hospitalisiert wurden – also Patienten mit einer «schweren» Infektion. Remdesivir zeigte für die Patienten im Vergleich zu Placebo keinen klinischen Benefit.

Ein- und Ausschlusskriterien lassen sich, wie bereits erwähnt, enger oder weiter definieren. Welche Variante – eng oder weit gefasst – geeigneter ist, bestimmt der jeweilige Kontext.

Beispiele (fiktiv):

Studie zur Untersuchung der Wirksamkeit – Reduktion von Exazerbationen – eines neuen Medikamentes bei Patienten mit chronisch, obstruktiver Lungenerkrankung (COPD).

Studie mit eng gefassten Einschluss-Kriterien:

Männer zwischen 18 und 35 Jahren mit COPD, einem FEV₁ zwischen 60 und 70 %, mindestens drei Exazerbationen der COPD im vergangenen Jahr und keine relevanten Begleiterkrankungen.

Es gibt wahrscheinlich wenige Patienten mit einer COPD, die diese Kriterien erfüllen. Wenn das neue Medikament bei diesen Patienten einen klinischen Benefit zeigt, dann ist das Ergebnis auf Patienten mit diesen Charakteristika übertragbar. Fragen, die sich aber stellen sind: z.B. Sind die Ergebnisse auch auf Frauen übertragbar? Sind sie auf über 70-Jährige übertragbar? Fragen, die sich nicht abschliessend beantworten lassen.

Studie mit breit gefassten Einschluss-Kriterien:

Frauen und Männer zwischen 50 und 90 Jahren mit COPD, einem FEV₁ < 60 % und mindestens einer Exazerbation in den letzten drei Jahren.

Die Ergebnisse einer solchen Studie sind einfacher auf ein grösseres Kollektiv

von Patientinnen und Patienten, die in einer Praxis behandelt werden, übertragbar.

Bei einer ersten Studie zur Untersuchung der Wirksamkeit eines neuen Medikaments werden die Ein- und Ausschlusskriterien in der Regel eng definiert. Dies dient dazu, die Wirksamkeit zunächst unter möglichst günstigen Bedingungen nachzuweisen – d.h. in einer Patientengruppe, bei der ein Effekt am wahrscheinlichsten zu beobachten ist. Erst nach erfolgreichem Wirksamkeitsnachweis werden die Kriterien in weiteren Studien oft breiter gefasst.

Aus Sicht vieler Ärztinnen und Ärzten ist dieses Vorgehen jedoch problematisch. In der täglichen Praxis erfüllen oft nur wenige ihrer Patientinnen und Patienten – selbst bei häufigen Erkrankungen – die eng gefassten Einschlusskriterien. Entsprechend schwierig ist es, die Frage zu beantworten welchen Effekt die untersuchte Therapie auf den individuellen Patienten hat.

Bei Studien zur Wirksamkeit einer Behandlung wird zwischen 'erklärenden' und 'pragmatischen' Studien unterschieden.

Erklärende Studien untersuchen, ob eine Therapie unter möglichst idealen und kontrollierten Bedingungen wirkt (wie zum Beispiel die Ein- und Ausschlusskriterien der 4S-Studie, siehe im Anhang). Dies bedingt enge Einschlusskriterien, eine engmaschige Betreuung der Teilnehmenden, Kontrollen der Medikamenteneinnahme sowie regelmässige Kontakte mit dem Studienpersonal.

Pragmatische Studien hingegen prüfen, wie gut eine Therapie unter Alltagsbedingungen funktioniert. Sie orientieren sich stärker an der klinischen Praxis: Die Einschlusskriterien sind meist breiter, die Betreuung ist weniger engmaschig, die Einnahmetreue wird oft nicht systematisch kontrolliert, und Begleitmedikationen sind oft erlaubt.

Entscheidend, ob man ein erklärendes oder pragmatisches Design wählt, ist die Fragestellung der Studie. Soll geklärt werden, ob eine Therapie unter kontrollierten Bedingungen (z.B. dass die Patienten die Medikamente regelmässig einnehmen) wirksam ist, eignet sich eine erklärende Studie. Soll hingegen untersucht werden, welchen Nutzen eine Therapie im klinischen Alltag hat, ist eine pragmatische Studie passender. Die Ergebnisse pragmatischer Studien bilden die Versorgungspraxis besser

ab, Aussagen zur eigentlichen Wirksamkeit einer Therapie müssen jedoch sehr vorsichtig interpretiert werden.

Was meint man mit «Übertragbarkeit der Ergebnisse»?

Aus klinischer Sicht stellt sich immer die Frage, ob die Ergebnisse einer Studie auf einen Patienten, der gerade vor der Ärztin im Sprechzimmer sitzt, übertragbar sind. Anders formuliert, was sind die erwartbaren Konsequenzen, wenn einem Patienten das Medikament, dessen Wirksamkeit in der Studie untersucht wurde, empfohlen wird. Zeigt sich also bei dem Patienten die gleich grosse Wirkung wie bei den Patientinnen und Patienten in der Studie? Nehmen wir im folgenden Beispiel der Einfachheit halber an, dass der betreffende Patient die Einschlusskriterien erfüllt hätte und somit an der Studie teilnehmen hätte können.

Beispiel: Der Patient ist ein 65-jähriger Mann mit einer Herzinsuffizienz NYHA III und einer linksventrikulären Auswurffraktion von 30 %. Er wird bereits mit einem Beta-Blocker, einem ACE-Hemmer sowie einem Mineralokortikoid-Rezeptor-Antagonisten behandelt. Es stellt sich die Frage, welchen zusätzlichen Nutzen eine Therapie mit Digitoxin bringen könnte.

Eine kürzlich publizierte Studie (6) ergab, dass Digitoxin im Vergleich zu Placebo wirksam ist. Digitoxin verringert das Risiko, innerhalb von drei Jahren zu sterben oder aufgrund der Herzinsuffizienz hospitalisiert zu werden. In die Studie eingeschlossen wurden Patienten mit einer symptomatischen chronischen Herz-insuffizienz (linksventrikuläre Auswurffraktion \leq 40% und NYHA Klasse III oder IV), die älter als 18 Jahre alt waren und eine evidenzbasierte Therapie für mindestens 6 Monate erhalten haben.

Die wenig differenzierte Konsequenz aus den Studienergebnissen wäre, Digitoxin aufgrund seiner statistisch signifikanten Überlegenheit gegenüber Placebo zu empfehlen. Eine Therapieentscheidung allein auf dieser Basis wird dem individuellen Patienten jedoch nicht gerecht. Ärztin und Patient benötigen zusätzliche Informationen, um Nutzen, Risiken und die bestehende Polypharmazie angemessen berücksichtigen zu können.

Die absolute Risikoreduktion innerhalb von drei Jahren zu versterben oder aufgrund der Herzinsuffizienz hospitalisiert zu werden betrug 4,6 % (der

Outcome trat bei 44,1 % in der Placebogruppe und bei 39,5 % in der Digitoxin-Gruppe ein).

Sieht man sich die Ergebnisse verschiedener Subgruppen an zeigt sich, dass Patientinnen und Patienten mit einer Herzfrequenz über 75/min deutlich stärker von Digitoxin profitierten als jene mit einer niedrigeren Herzfrequenz. Ebenso profitieren Patienten mit einem systolischen Blutdruck unter 120 mmHg, während sich bei denen mit einem systolischen Blutdruck über 120 mmHg kein Benefit zeigt.

Für die therapeutische Entscheidung sind jedoch nicht nur die potenziellen Vorteile, sondern auch die Nebenwirkungen relevant. In der Digitoxin-Gruppe traten schwere Nebenwirkungen häufiger auf als unter Placebo, insbesondere Kammerflimmern und ventrikuläre Tachykardien (2,7 % versus 1,3 %).

Endpunkt: Sterblichkeit oder Hospitalisation

	Hazard Ratio (95% Konfidenzintervall) ³	Absolute Risikoreduktion
Alle Pat.	0.82 (0.69-0.98)	4.6%
Pat. mit Herzfrequenz < 75/min	0.92 (0.72-1.18)	3.2 %
Pat. mit Herzfrequenz ≥ 75/min	0.63 (0.48-0.83)	8.2 %
Pat. mit Blutdruck ≤ 120 mm Hg	0.61 (0.48-0.79)	9.7 %
Pat. mit Blutdruck > 120 mm Hg	1.03 (0.78-1.36)	+ 0.6 %

Gravierende Nebenwirkungen

	Digitoxin-Gruppe	Placebo-Gruppe
Alle Patienten	4.7%	2.8%

Diese Informationen – unterschiedliche Wirksamkeit in Subgruppen, teilweise lebensgefährliche Nebenwirkungen – sind relevant um dem Patienten die Einnahme von Digitoxin zu empfehlen oder ihm eher davon abzuraten.

Die externe Validität einer Studie beschreibt, wie gut sich deren Ergebnisse auf Patientinnen und Patienten außerhalb der Studie übertragen lassen. Entscheidend

³ Siehe den Artikel «Darstellung und Interpretation von Ergebnissen randomisierter Studien»

dafür sind vor allem die Ein- und Ausschlusskriterien: Je enger diese gefasst sind, desto eingeschränkter ist meist die Übertragbarkeit auf die klinische Praxis. Beispiele mit Aspirin und Remdesivir zeigen, dass unterschiedliche Studienpopulationen zu unterschiedlichen Ergebnissen führen können. Neben der Grösse der Wirksamkeit sind auch Nebenwirkungen und Subgruppenergebnisse zu berücksichtigen, um zu beurteilen, ob ein Studienergebnis auch für einen konkreten Patienten zutrifft.

Anhang:

Ein und Ausschlusskriterien der 4 S Studie (7)

Einschlusskriterien:

Frauen und Männer 35 bis 70 Jahre alt

Angina pectoris oder Myokardinfarkt in der Anamnese

Serumcholesterin 5.5 bis 8.0 mmol/L und Serum-Triglyceride ≤ 2.5 mmol/L nach zwei Wochen Placebo-Einnahme

Ausschlusskriterien

Prämenopausale Frauen, sekundäre Hypercholesterinämie, instabile oder Prinzmetal Angina, Myokardinfarkt in den vergangenen 6 Monaten vor Einschluss in die Studie, antiarrhythmische Therapie, Herzinsuffizienz (behandelt mit Digitalis, Diuretika oder Vasodilatoren), hämodynamisch relevante Klappenveränderungen, Schlaganfall in der Anamnese, eingeschränkte Leberfunktion, partieller ilealer Bypass, Drogen- oder Alkoholabus in der Anamnese, geringe mentale Funktion (?), andere gravierende Erkrankungen, Behandlung mit anderen Studienmedikamenten, Hypersensitivität gegen HMG-CoA Reduktase Inhibitoren (Statin)

Referenzen

1. Steering Committee of the Physicians' Health Study Research G. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1989;321(3):129–35.
2. Gaziano JM, Brotons C, Coppolecchia R, Cricelli C, Darius H, Gorelick PB, et al. Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial. *Lancet.* 2018;392(10152):1036–46.
3. Group ASC, Bowman L, Mafham M, Wallendszus K, Stevens W, Buck G, et al. Effects of Aspirin for Primary Prevention in Persons with Diabetes Mellitus. *N Engl J Med.* 2018;379(16):1529–39.
4. Gottlieb RL, Vaca CE, Paredes R, Mera J, Webb BJ, Perez G, et al. Early Remdesivir to Prevent Progression to Severe Covid-19 in Outpatients. *N Engl J Med.* 2022;386(4):305–15.

5. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet*. 2020;395(10236):1569–78.
6. Bavendiek U, Grosshennig A, Schwab J, Berliner D, Rieth A, Maier LS, et al. Digitoxin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2025;393(12):1155–65.
7. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994;344(8934):1383–9.