

## Darstellung und Interpretation von Ergebnissen randomisierter Studien

Prof. em. Dr. med. Johann Steurer

Das Verfassen dieses Artikels wäre ohne die grosse Unterstützung von Prof. Ulrike Held und Claudia Lourenço Rodrigues, MSc. vom Institut für Epidemiologie, Biostatistik und Prävention der Universität Zürich nicht möglich gewesen.

Wie im Artikel – *Prognose – der weitere Verlauf einer Krankheit* – beschrieben, wird die Wirksamkeit von Therapien – etwa von Medikamenten, Operationen oder psychotherapeutischen Verfahren – wenn immer möglich mit randomisierten Studien (RCTs) untersucht.

In den Ergebnissen oder Schlussfolgerungen dieser Studien heisst es häufig, dass das untersuchte Medikament oder die Therapie „besser“ sei als ein Placebo oder eine andere Behandlung (z.B. eine bereits etablierte Therapie). Diese Aussage ist zwar formal korrekt, bleibt jedoch auf einer qualitativen Ebene und ist für Ärztinnen und Patienten wenig informativ. Sowohl Ärztinnen und Ärzte als auch Patientinnen und Patienten wünschen in der Regel konkretere Informationen, insbesondere zur Effektstärke (Wirksamkeit) einer Behandlung sowie zu möglichen Nebenwirkungen. Entscheidend sind daher quantitative Antworten auf folgende Fragen:

- Wie gross ist der Unterschied im klinischen Outcome zwischen der Behandlungs- und der Kontrollgruppe?
- Wie präzise ist die Schätzung dieses Unterschieds, ausgedrückt mit dem Konfidenzintervall?

Nur wenn diese Zahlen angegeben werden – z. B. die absolute Risikoreduktion (ARR), die relative Risikoreduktion (RRR) und die Number Needed to Treat (NNT) (1), lässt sich die tatsächliche klinische Relevanz einer Therapie beurteilen.

*Konkretes Beispiel zur Wirksamkeit von Antibiotika bei Harnwegsinfekten.*

*Harnwegsinfekte treten häufig auf, besonders bei Frauen. Eine antibiotische Behandlung kann den Krankheitsverlauf verkürzen. Da jedoch Resistenzen gegen Antibiotika zunehmen, müssen Behandlungsempfehlungen regelmäßig angepasst werden.*

*Bereits 2011 empfahlen einige Leitlinien Fosfomycin oder Nitrofurantoin als Mittel der ersten Wahl (2). Auch 2024 sind beide Substanzen in vielen Empfehlungen weiterhin als bevorzugte Optionen aufgeführt (3).*

*Eine randomisierte Studie (4) verglich die Wirksamkeit dieser beiden Medikamente bei akuten, unkomplizierten Harnwegsinfekten:*

- *258 Frauen erhielten eine Einmaldosis Fosfomycin (3 g),  
255 Frauen nahmen Nitrofurantoin 100 mg dreimal täglich über fünf Tage ein.*

*Outcome (wie quantifiziert man den Erfolg der Behandlung?):*

*Vollständige Symptomfreiheit vier Wochen nach Behandlungsbeginn – keine Beschwerden wie Brennen beim Wasserlassen, Pollakisurie, Harndrang oder suprapubische Schmerzen.*

*Im Abstract berichten die Studienautoren, dass nach vier Wochen Nitrofurantoin signifikant häufiger zur Abheilung führte als Fosfomycin.*

*Doch reicht diese Information aus, um Nitrofurantoin generell zu bevorzugen?*

*Nicht unbedingt. Für eine informierte Entscheidung – sowohl aus ärztlicher als auch aus Patientinnensicht – sind weitere Details wichtig:*

- *Wie grossist der Unterschied im Behandlungserfolg?*
- *Gibt es Nebenwirkungen?*
- *Wie präzise ist die Schätzung des Unterschieds? (95 % Konfidenzintervall (5))*
- *und der p-Wert, der aber vollkommen überflüssig und manchmal verwirrend ist, wenn man das 95% Konfidenzintervall kennt (p-Wert) (6).*

*Ergebnisse im Detail:*

- *Beschwerdefrei nach 4 Wochen:*
  - *Nitrofurantoin-Gruppe: 70 %*
  - *Fosfomycin-Gruppe: 58 %*
- *Therapieversagen (nicht beschwerdefrei):*
  - *Nitrofurantoin: 30 %*
  - *Fosfomycin: 42 %*
- *Nebenwirkungen traten in beiden Gruppen bei etwa 7 % auf; in erster Linie gastrointestinale Nebenwirkungen (Übelkeit, Durchfall).*

*Schwerwiegende Nebenwirkungen traten in keiner der beiden Gruppen auf.*

*Quantitative Auswertung der Effektstärke:*

- 1. Absolute Risikoreduktion (ARR):  $42\% - 30\% = 12\%$*
- 2. Relative Risikoreduktion (RRR):  $12\% \div 42\% = 29\%$*
- 3. Number Needed to Treat (NNT):  $100 \div 12 = 8$ . Man muss 8 Frauen mit Nitrofurantoin statt Fosfomycin behandeln, um eine zusätzliche Heilung zu erreichen.*
- 4. 95 % Konfidenzintervall – Präzision der Schätzung – für die absolute Risikoreduktion von 12 % (95% Konfidenzintervall 4 % bis 21 %). Der Unterschied im Outcome ist statistisch signifikant (auf dem 5% Signifikanzniveau), da das Konfidenzintervall den Wert Null nicht enthält. Würde man die Studie viele Male wiederholen, dann läge der wahre Unterschied mit sehr hoher Wahrscheinlichkeit innerhalb des 95%-Konfidenzintervalls.*
- 5. p-Wert = 0.004. Der p-Wert gibt die Wahrscheinlichkeit an, dass das beobachtete Ergebnis oder ein extremeres unter der Annahme der Nullhypothese (d.h. kein Unterschied zwischen Fosfomycin und Nitrofurantoin) zustande gekommen ist.*

## **Fazit**

Die Studie zeigt, dass Nitrofurantoin bei unkomplizierten Harnwegsinfektionen statistisch signifikant wirksamer ist als eine einmalige Fosfomycin-Gabe. Die absolute Differenz im Behandlungserfolg liegt bei 12 %, was in der Praxis relevant ist. Die Nebenwirkungen sind gemäss dieser Studie in beiden Gruppen sehr ähnlich und gleich häufig. Die Kosten sind in diesem Fall mit ca. 13 Franken pro Therapie vergleichbar und daher kein entscheidendes Kriterium für die eine oder andere Behandlung.

## **Absolute versus relative Risikoreduktion**

Die Wirksamkeit von Medikamenten oder anderen therapeutischen Maßnahmen wird häufig in Form der relativen Risikoreduktion (RRR) dargestellt – insbesondere in

der Pharmawerbung. Diese Darstellung kann den Eindruck einer besonders hohen Effektivität erwecken, obwohl der tatsächliche Nutzen oft wesentlich geringer ausfällt, wenn man die absolute Risikoreduktion (ARR) berücksichtigt.

*Ein Beispiel verdeutlicht diesen Unterschied: 1996 wurde eine randomisierte Studie zur Wirksamkeit von Alendronat, einem Bisphosphonat zur Steigerung der Knochendichte bei Osteoporose, veröffentlicht (7). In einer Broschüre wurde damals verkündet, dass mit dem Medikament das Risiko für Wirbelkörperfrakturen um 54 % gesenkt werde – eine Angabe, die auf der relativen Risikoreduktion basiert und sehr eindrucksvoll klingt.*

*Ein Blick auf die absoluten Zahlen zeigt jedoch ein anderes Bild: In der Placebo-Gruppe kam es bei 5.0 % der Teilnehmerinnen zu einer Wirbelkörperfraktur, in der Alendronat-Gruppe bei 2.3 %. Das entspricht einer absoluten Risikoreduktion von lediglich 2.7 %, oder einer Number Needed to Treat (NNT) von 37 – das heißt, es müssen 37 Frauen behandelt werden, um eine Fraktur zu verhindern. (Die Beobachtungsdauer betrug 3 Jahre, also muss man 37 Frauen 3 Jahre behandeln, um eine Wirbelkörperfraktur zu verhindern).*

Dieses Beispiel unterstreicht die Bedeutung, bei der Bewertung der Wirksamkeit medizinischer Interventionen nicht nur die relative, sondern insbesondere auch die absolute Risikoreduktion zu berücksichtigen.

### **Hazard ratio**

In vielen randomisierten Studien wird das Ergebnis in der Form der «Hazard Ratio» angegeben.

Beispiel: In einer randomisierten Studie (Sprint-Studie) wurde bei Patienten mit einer Hypertonie der Unterschied im Outcome zwischen den Gruppen mit einer intensiven (systolischer Blutdruck < 120 mmHg) und einer Standard-Blutdrucksenkung (systolischer Blutdruck < 140 mmHg) untersucht (8). Der Outcome in der Studie war ein zusammengesetzter ('composite') Outcome, bestehend aus Herzinfarkt oder akutem koronarem Syndrom oder Schlaganfall oder noch weiteren kardiovaskulären Ereignissen. Die mediane Beobachtungsdauer betrug 3.26 Jahre.

	Intensive Behandlung		Standard- Behandlung		Hazard Ratio	p-Wert
Outcome	%	% pro Jahr	%	% pro Jahr		
Primärer Outcome	5.2	1.65	6.8	2.19	0.75 (0.64 – 0.89)	< 0.001

Die absolute Risikoreduktion:  $6.8 - 5.2 = 1.6 \%$

Die relative Risikoreduktion:  $1.6 \div 6.8 = 24 \%$

Die Hazard Ratio ist der Vergleich der Rate des primären Outcomes in der intensiv behandelten Gruppe mit der Rate in der Standard-Behandlungsgruppe während der Beobachtungsphase (Studiendauer); die Hazard ratio ist 0.75; das heisst die Wahrscheinlichkeit des Eintretens des Outcomes ist in der intensiv behandelten Gruppe, im Vergleich zur Standard-Behandlungsgruppe, um 25 % reduziert.

### **Welche zusätzliche Information enthält die Hazard Ratio im Vergleich zur relativen Risikoreduktion?**

Für die Berechnung der Hazard Ratio werden die sog. *Ereigniszeiten* betrachtet, also die Zeit zwischen der Randomisierung und dem Auftreten des Ereignisses (z.B. des composite-Outcomes in der Sprint Studie; oder Überleben bei Patienten mit einem Tumor).

Falls kein Ereignis eintritt, kann die Hazard Ratio mit sog. Zensierungen umgehen. Das bedeutet, dass auch die Ereigniszeit von Patienten berücksichtigt werden kann, die die Studie verlassen (vor der vorgesehenen Studiendauer/Nachbeobachtungszeit nach Studieneinschluss).

#### *Beispiel*

*Ein Patient wird in eine Studie eingeschlossen, in der das Überleben als primärer Endpunkt definiert ist. Er erscheint zur Kontrolle nach sechs Monaten, jedoch nicht nach 12 Monaten und gilt als Drop-out. Für die Berechnung der Risikoreduktion (RRR) nach zwölf Monaten kann dieser Patient nicht berücksichtigt werden, da unbekannt ist, ob er zu diesem Zeitpunkt noch lebt. Wird hingegen die Hazard Ratio (HR) berechnet, geht*

*die Information in die Analyse ein, dass der Patient zum Zeitpunkt des letzten Kontakts (sechs Monate) noch am Leben war.*

Daraus folgt:

Bei Studien mit keinen oder nur wenigen Drop-outs unterscheiden sich RRR und HR nur geringfügig. Bei vielen Drop-outs hingegen können die Werte deutlich auseinanderliegen.

Im SPRINT Trial war die Anzahl der Zensierungen gering, denn die Autoren schrieben in der Fallzahlplanung «anticipated a loss to follow-up of 2% per year». Daher ist der Unterschied zwischen der HR (0.75) und der RRR (24 %; oder  $1 - RRR = 0.76$ ) sehr gering.

*Ein Beispiel mit höherer Drop-Out Rate:*

*In einem hypothetischen Beispiel werden 228 Patienten mit Lungenkrebs ab Beginn der Therapie über die Zeit untersucht. Hierbei soll das Überleben in zwei Gruppen verglichen werden. Es wird in der folgenden Tabelle dargestellt, wie für Gruppe A versus B, die Hazard Ratio, und das Relative Risiko bei 2 Jahren Follow-up geschätzt werden. Zwei unterschiedliche Prozentsätze an Zensierungen wurden untersucht.*

<i>Zensierung («drop out» Rate)</i>	<i>Hazard Ratio (95 % Konfidenzintervall)</i>	<i>Relatives Risiko (RR)</i>
<i>30 %</i>	<i>0.56 (0.4-0.78)</i>	<i>0.95 (0.86-1.05)</i>
<i>51 %</i>	<i>0.6 (0.41-0.89)</i>	<i>1.04 (0.82-1.31)</i>

Interpretation: Die Hazard Ratio wird durch die deutliche Zunahme der Zensierungen kaum beeinflusst. Der Wert für das Relative Risiko unterscheidet sich deutlich von der Hazard Ratio und verändert sich bei höherem Prozentsatz von Zensierungen noch stärker. Wenn eine Entscheidung zwischen A und B getroffen werden muss (aufgrund der in der obigen Tabelle dargestellten HR-Werte), dann ist A gegenüber B zu bevorzugen (der Unterschied ist statistisch signifikant). Bei den RR Werten sind die Unterschiede zwischen A und B –

infolge der grossen Anzahl an Zensierungen (drop-outs) – nicht signifikant. Mit zunehmender Zahl an Zensierungen wird die Schätzung des RR verzerrt und unpräziser.

## Referenzen

1. Sedgwick P. Number needed to treat I. *BMJ*. 2011;342.
2. Gupta K, Hooton TM, Naber KG, Wullt B, Colgan R, Miller LG, et al. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis*. 2011;52(5):e103–20.
3. Nelson Z, Aslan AT, Beahm NP, Blyth M, Cappiello M, Casaus D, et al. Guidelines for the Prevention, Diagnosis, and Management of Urinary Tract Infections in Pediatrics and Adults: A WikiGuidelines Group Consensus Statement. *JAMA Netw Open*. 2024;7(11):e2444495.
4. Huttner A, Kowalczyk A, Turjeman A, Babich T, Brossier C, Eliakim-Raz N, et al. Effect of 5-Day Nitrofurantoin vs Single-Dose Fosfomycin on Clinical Resolution of Uncomplicated Lower Urinary Tract Infection in Women: A Randomized Clinical Trial. *JAMA*. 2018;319(17):1781–9.
5. Sedgwick P. Confidence intervals, P values, and statistical significance. *BMJ*. 2015;350:h1113.
6. Sedgwick P. Understanding P values. *BMJ*. 2014;349:g4550.
7. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, et al. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *The Lancet*. 1996;348(9041):1535–41.
8. SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015;373(22):2103–16.