DER ARZNEIMITTELBRIEF

Jg. 38, S. 23; Ausgabe 03 / 2004

Leserbrief: Zur Aussagekraft der Number needed to treat

Prof. Dr. J. Windeler, Essen, schreibt zu unserer Leserbrief-Beantwortung (AMB 2003, **37**, 63): >> David Sackett handelte zweifellos in bester Absicht und aus fundierter praxisorientierter Überlegung heraus, als er das Konzept der Number Needed to Treat (NNT) als Reziprok der absoluten Risikodifferenz vorschlug. Es stellt sich jedoch die Frage, ob er diesen Vorschlag auch gemacht hätte, wenn er den verbreiteten Mißbrauch des Begriffs und des Konzepts vorhergesehen hätte. So sehr NNT noch für die quantitative Bewertung der Ergebnisse einer einzelnen Studie nützlich sein können, so problematisch ist ihre Verwendung in der vergleichenden Bewertung von Studien oder gar von Therapien. Die auftretenden Probleme bringen NNT in diesem Zusammenhang hart an die Grenze der Unbrauchbarkeit.

Zum vorgelegten Vergleich von NNT verschiedener medikamentöser und nicht-medikamentöser Maßnahmen zur kardiovaskulären Sekundärprävention seien folgende Bemerkungen gemacht, die sich aber generell auf die Verwendung von NNT in solchen Problemstellungen beziehen:

a) Konfidenzintervalle: Die NNT ist genau so wie jede andere empirische Maßzahl ein Schätzer mit einer mehr oder weniger ausgeprägten Ungenauigkeit. Zu dieser Maßzahl gehört daher unbedingt ein Konfidenzintervall, welches sich im Übrigen relativ einfach berechnen läßt. Die Tab. 1 zeigt die Konfidenzintervalle für die Ergebnisse der benannten Studien, jeweils für die gesamte Follow-up-Periode und für den kombinierten Endpunkt kardiovaskuläre Todesfälle oder nichttödlicher Myokardinfarkt: Die Lifestyle-Studie, SCRIP sowie die Stress-Management-Studie wurden nicht berücksichtigt.

Man erkennt die teilweise große Unsicherheit in den Angaben auch bei großen Studien. Dies macht deutlich, daß mit der Nennung und Interpretation von Zahlen wie 153 ein falscher Eindruck erweckt wird. NNT sollten eher in Klassen von 10er-Potenzen eingeordnet werden.

b) Für die Interpretation einer Zahl, die beschreibt, wie viele Patienten für einen zusätzlichen Erfolg behandelt werden müssen, ist Voraussetzung, daß diese Erfolgsaussicht überhaupt besteht. Mindestvoraussetzung dafür ist ein signifikant positives Studienergebnis. Aus diesem Grund ist die Berechnung von NNT aus Studien mit nicht-signifikanten Ergebnissen wie der Lifestyle Heart Trial mit 40 (!) Patienten und 2 bzw. 4 Zielereignissen oder der Stress-Management-Studie mit 0 und 3 Zielereignissen grundsätzlich nicht sinnvoll (und schon überhaupt nicht ohne Konfidenzintervall, was bei einer Betrachtung des Konfidenzintervalls auch sofort klar geworden wäre).

c) Es ist zwar gut verständlich, daß in der Zielsetzung, die Ergebnisse verschiedener Studien miteinander vergleichen zu können, die NNT auf ein Jahr standardisiert werden. Das Vorgehen ist jedoch problematisch, da für eine sinnvolle Interpretation Voraussetzungen bezüglich der „Kinetik"

Diese Publikation ist urheberrechtlich geschützt. Vervielfältigung sowie Einspeicherung und Verarbeitung in elektronischen Systemen ist nur mit Genehmigung der Herausgeber des ARZNEIMITTELBRIEFS gestattet.

der kumulativen Ereignisraten erfüllt sein müssen. Offensichtlich wird dies, wenn zwei Studien zu einer identischen Therapie gegenübergestellt werden, bei der der Effekt erst mit Latenz (z.B. nach 3 Jahren) einsetzt. Auf ein Jahr standardisierte Ergebnisse einer zweijährigen Studie werden dann völlig anders ausfallen als die einer fünfjährigen Studie. Der Unterschied würde vordergründig als Effekt der Therapie interpretiert; tatsächlich ist er aber weder in der Therapie noch in unterschiedlichen Charakteristika der Patienten zu suchen, sondern ausschließlich in der unterschiedlichen Dauer der Studien. Der Einjahresbezug setzt für eine sinnvolle Interpretation voraus, daß die absolute Risikodifferenz und damit der vertikale Abstand von Überlebenskurven über die Zeit linear verläuft. Welche Relevanz eventuelle Abweichungen für eine einzelne Studie erlangen, ist nicht allgemein zu beantworten, eine größere Zurückhaltung bei der Verwendung solcher Standardisierungen aber unbedingt anzuraten.

d) Das gravierendste Problem eines Vergleichs von NNT aber wird bereits aus der Bemerkung eines Leserbriefschreibers (AMB 2003, **37**, 15b) deutlich, der meinte „NNT im dreistelligen Bereich" als „Witzveranstaltung" apostrophieren zu können. Es gerät bei der verbreiteten Wertschätzung von NNT offenbar völlig aus dem Blick, daß es sich hier um *absolute* Differenzen handelt. Dieses bedeutet, daß NNT vom Baseline-Risiko (abgebildet durch die Ereignisrate in der Kontroll-Gruppe) abhängig sind. Liegt, z.B. in der Primärprävention, das Risiko für ein bestimmtes Ereignis innerhalb von 5 Jahren unterhalb von 1%, so kann eine Intervention niemals eine niedrigere als dreistellige NNT erreichen, selbst dann nicht, wenn die Intervention das in Frage stehende Zielereignis komplett verhindert. Damit wird aber auch deutlich, daß ein Vergleich von NNT zwischen Studien nur dann möglich ist, wenn die Studien aufgrund ihrer Patientenauswahl und ihrem sonstigen Kontext gleiche oder wenigstens vergleichbare Baseline-Risiken aufweisen. Diese gehören zur Nennung von NNT *unabdingbar* dazu. Daß es sich zudem um Studien gleicher und ausreichender Qualität handeln muß, sei nur der Vollständigkeit halber erwähnt. Trägt man für die genannten Studien die Gesamt-NNT gegen die Baselinerisiken auf, so sieht man im Rahmen zu erwartender Schwankungen genau den erwarteten Zusammenhang: je höher das Ausgangsrisiko, desto geringer die NNT (Abb. 1).

NNT sollten nur zusammen mit dem zugehörigen Baseline-Risiko und einem Konfidenzintervall angegeben werden. Daß mit dem Ein-Jahres-Bezug nicht grobe Fehleinschätzungen befördert werden, ist jeweils zu begründen. Nur auf dieser Basis ist ein – zweifellos wenig formaler – Vergleich zwischen Studien möglich. In diesem Zusammenhang sei daran erinnert, daß die Betrachtung absoluter Differenzen *eine* Möglichkeit der Darstellung ist, *in Ergänzung* zu relativen Unterschiedsmaßen (Hazard-Ratio, Relatives Risiko, Odds-Ratio).

Antwort: >> Der Brief von Prof. Dr. Windeler trägt sehr zum besseren Verständnis der statistischen Methode bei, die Bedeutung einer Prophylaxe oder Therapie durch die Anzahl der Patienten zu beschreiben, die behandelt werden müssen, um einem Patienten ein klinisches Ereignis zu ersparen (Number needed to treat = NNT). Dazu müssen in der Regel Konfidenzintervalle angegeben werden, zumindest für die Zahlen, aus denen die NNT errechnet wurde. Für nicht-pharmakologische Interventionsstudien, in denen Änderungen des Lebensstils untersucht werden, stehen leider wenig Drittmittel zur Verfügung. Daher ist die Fallzahl oft klein. Die Ergebnisse von Meilensteinstudien aus diesem Bereich müssen aber – auch wenn sie klein sind – in ihrer Aussagekraft eingeschätzt werden,

Diese Publikation ist urheberrechtlich geschützt. Vervielfältigung sowie Einspeicherung und Verarbeitung in elektronischen Systemen ist nur mit Genehmigung der Herausgeber des ARZNEIMITTELBRIEFS gestattet.

nicht zuletzt deshalb, weil die Bedeutung von Lebensstilmodifikationen zwar allgemein anerkannt, aber nicht umgesetzt werden. Die größeren Studien in diesem Bereich genügen auch höheren methodischen Ansprüchen (1, 2).

Die Abb. 1 zeigt klar die Abhängigkeit der NNT vom Risiko der Kontroll-Gruppe und damit eine Grunderfahrung ärztlicher Praxis: je höher das Risiko, um so effektiver kann geholfen werden, d.h. um so niedriger ist die Zahl der Patienten, die behandelt werden müssen, um einem ein Ereignis zu ersparen. Es ist in unseren Augen die Stärke des NNT-Konzepts, daß es die statistische Relevanz einer Maßnahme abbilden kann. Eine teure Therapie ist z.B. bei hohem Risiko und niedriger NNT vertretbar, bei niedrigem Risiko und hoher NNT vielleicht nicht. Andererseits kann eine harmlose „Intervention" vielleicht mehreren (hundert) Patienten über Jahre zugemutet werden, um einem ein Ereignis zu ersparen. Nach unserer Meinung erleichtert das NNT-Konzept somit die Vermittlung pharmako-ökonomischer Zusammenhänge. Natürlich begründet aber der Sachverhalt, der in der Abbildung dargestellt ist, klar die Forderung, nur solche Studienergebnisse mit dieser statistischen Methode zu vergleichen, in deren Kontroll-Gruppen das Risiko etwa gleich ist. Überhaupt können Studienergebnisse nur dann übertragen werden, z.B. auch auf eine individuelle ärztliche Entscheidung, wenn die klinische Situation der Studienpopulation dem individuellen Patienten ähnlich ist.

Für den behandelnden Arzt ist es wichtig, Studienergebnisse miteinander vergleichen zu können. Eine NNT mit Zeitbezug (z.B. pro Jahr oder pro 2 Jahre) kann dabei wesentlich helfen, vorausgesetzt Konfidenzintervalle, Signifikanzen, Fehlinterpretationen durch Zeitbezug und Abhängigkeit der NNT vom Basisrisiko werden beachtet. Der bloße Quotient aus Risiko in der Kontroll-Gruppe und Risiko in der Interventionsgruppe (Hazard ratio) sagt wenig aus über die Bedeutung von Studienergebnissen und ist daher nicht sehr hilfreich bei der Therapieplanung für Gruppen.

Literatur

1. de Lorgeril, M., et al. (Lyon Diet Heart Study): Circulation 1999, **99**, 779; s.a AMB 1999, **33**, 74 und 2003, **37**, 8 und 63.
2. Singh, R.B., et al. (Indo-Mediterranean Diet Heart Study): Lancet 2002, **360**, 1455; s.a. AMB 2003, **37**, 8 und 63.

Tabelle 1
NNT mit Konfidenzintervallen verschiedener Studien

Studie	NNT	n	95%-KI
Lyon Diet Heart	7	433	[5; 12]
Indo-Med. Diet Heart	14	1000	[9; 30]
HPS	33	20536	[26; 45]
HOPE	45	9297	[29; 108]
PROSPER	48	5804	[27; 200]
GISSI	94	5664	[49; 1357]

Diese Publikation ist urheberrechtlich geschützt. Vervielfältigung sowie Einspeicherung und Verarbeitung in elektronischen Systemen ist nur mit Genehmigung der Herausgeber des ARZNEIMITTELBRIEFS gestattet.