

Klinische Fachliteratur kritisch lesen

Teil II: Randomisierte Studien, Meta-Analysen und Publikationsprobleme

M. F. Fey^a, A. Bühler^b

Einleitung

Fallberichte und einzelarmige klinische Studien weisen Grenzen in ihrer Aussagekraft auf; worauf wir im vorangegangenen Artikel mit Beispielen hingewiesen haben. Zur Etablierung neuer Therapierichtlinien genügen sie nicht, wenn kleine (aber klinisch potentiell relevante) Unterschiede in der Wirkung neuer *versus* herkömmlicher Therapien sauber belegt werden müssen. Randomisierte Studien sind in diesen Fällen ein wertvoller, mitunter der einzige Weg, die nötige wissenschaftliche Evidenz zu erwerben. So ist es denn wichtig, allfällige Fussangeln und Fallstricke bei der Interpretation dieser Studien zu kennen. Wir präsentieren einige Denkanstösse zu diesem Thema, gewürzt mit Beispielen aus der Onkologie.

Randomisierte Studien

In randomisierten Studien werden die Resultate einer neuen Behandlung in direktem Vergleich zum Kontrollarm geprüft, und die Verläufe (z.B. Verlängerung des Überlebens, verminderte oder verstärkte Toxizität) können damit direkt auf die Therapie zurückgeführt werden. Die Randomisation, d.h. die zufällige Zuteilung von Patienten vor Therapiebeginn zu einem «alten» Therapie- oder Kontrollarm (bisheriger Standard) oder einem «neuen» Arm (neue klinisch «experimentelle» Therapie) löst damit einige der im ersten Artikel erwähnten Analyse- und Interpretationsprobleme, die andere klinische Studienanlagen bieten. Sie sichert zu, dass sich die zu vergleichenden Patientengruppen nur hinsichtlich Therapie unterscheiden, dass aber andere Faktoren, die den Verlauf mitbestimmen, in beiden Gruppen gleich verteilt sind. Dies betrifft bekannte prognostische Faktoren, aber auch unbekannte prognostische Faktoren. Erstere könnten bei der Bildung historischer Kontrollgruppen allenfalls berücksichtigt werden, letztere aber auf keinen Fall. So sind zum Beispiel molekular-

biologische Faktoren (inaktive Tumor-Suppressor-Gene, aktivierte Onkogene und andere mehr) entweder in der Routine-Diagnostik nicht ohne weiteres erfassbar oder zum Zeitpunkt einer Studie (noch) gar nicht bekannt.

Die Durchführung randomisierter Studien sollte gewährleisten, dass Patienten, die eine etablierte Therapie erhalten, gleich sorgfältig behandelt und dokumentiert werden wie diejenigen, die unter eine neue Behandlung zu stehen kommen. Patienten werden unter einer neuen (weniger vertrauten) Therapie oft besonders sorgfältig betreut, und die bessere Gesamtbetreuung mag einen besseren Verlauf, unabhängig von der eigentlichen Behandlung, mit sich bringen. Trotz dieser Vorteile sind auch randomisierte Studien nicht vor Problemen gefeit.

Endpunkte randomisierter Studien. Endpunkte sind die Masseinheiten, mit denen ein Therapieeffekt gemessen werden soll. Bei der Lektüre eines Artikels mit Daten aus einer randomisierten Studie stellt sich die Frage, ob die gewählten Endpunkte klinisch sinnvoll seien (Tab. 1). Bei der adjuvanten medikamentösen Therapie des Mammakarzinoms, welche die Prognose der Patientinnen in einem kurativen Konzept verbessern soll, sind Endpunkte wie Rezidiv-freies Überleben und Gesamtüberleben besonders wichtig. Bei einer Studie zur palliativen Therapie einer metastasierenden Tumorkrankheit sind Parameter wie Lebensqualität oft gewichtiger als das Gesamtüberleben. Die Verbesserung der Remissionsrate bei Patienten mit einem metastasierenden Karzinom ist klinisch nur sinnvoll, wenn sie assoziiert ist mit Lebensverlängerung oder der Besserung subjektiver Parameter. Falls die Therapie lediglich zur Rückbildung asymptomatischer Metastasen im Röntgenbild führt, aber um den Preis erheblich höherer Toxizität und bei unverändertem Gesamtverlauf, so ist ein echter therapeutischer Fortschritt fraglich.

Der Standard- oder Kontroll-Arm in randomisierten Studien. Da der Standard- oder Kontrollarm als Referenz für die Beurteilung einer neuen Therapie dient, sind seine Resultate mit besonderer Sorgfalt zu betrachten.

- Entspricht die Therapie im Standard-Arm dem heutigen Standard? Man ziehe publizierte Richtlinien («guidelines»), und aktuelle Übersichtsartikel (mit Vorliebe von anderen Autoren) zu Rat.
- Entsprechen die Resultate des Standard-Arms den Erwartungen? Fallen sie schlechter aus, als gemeinhin erwartet werden darf, so kommt dies dem «experimentellen Arm» zugute, auch wenn die neue Therapie im Grunde genommen keine wesentliche Verbesserung darstellt. Sind die Resultate

^a Institut für medizinische Onkologie, Inselspital Bern

^b Innere Medizin FMH, Novazzano

Korrespondenz:
Prof. M. Fey
Institut für medizinische Onkologie
Inselspital Bern
CH-3010 Bern

Tabelle 1. Beispiele für die Wahl von Endpunkten klinischer Studien.

	Primäre d.h. wichtigste Endpunkte	Sekundäre Endpunkte
Kuratives Konzept		
Adjuvante Chemotherapie beim Mammakarzinom	Gesamtüberleben	Rezidivfreies Überleben
Palliatives Konzept		
Chemotherapie beim metastasierenden Lungenkarzinom	Lebensqualität, progressionsfreies Überleben	Remissionsrate Gesamtüberleben
Supportive Massnahme		
Hämatopoietische Wachstumsfaktoren nach Chemotherapie	Reduktion des Infektrisikos bei Neutropenie und Reduktion des Antibiotika-Verbrauchs	Besserung hämatologischer Laborwerte, Überleben

- des Standard-Arms über Erwarten gut, so ist ein Unterschied zwischen Kontrolle und experimenteller Therapie unter Umständen nicht belegbar.
- Wurden die Behandlung und der klinische Verlauf im Standard-Arm ähnlich stringent kontrolliert wie im experimentellen Arm (Anzahl der klinischen Visiten, Frequenz der Labor-Untersuchungen, der radiologischen Kontrollen)?

Die Schweizerische Arbeitsgemeinschaft für Klinische Krebsforschung (SAKK) führte über Jahre Studien zum Thema der adjuvanten Chemotherapie beim radikal operierten kolorektalen Karzinom durch. Operierte Patienten wurden entweder beobachtet (Standard-Arm) oder zusätzlich mit Chemotherapie behandelt (experimenteller Arm). In der zweiten Studien-Generation fielen die Resultate der rein chirurgischen Behandlung im Standard-Arm wesentlich besser aus als in der ersten Studien-Generation. Ein zusätzlicher Effekt einer adjuvanten Chemotherapie war somit schwieriger zu belegen (U. Laffer, Biel – persönliche Mitteilung). In einer randomisierten Studie wurde der Stellenwert der regionalen Chemotherapie bei Lebermetastasen kolorektaler Karzinome überprüft [1]. Patienten im experimentellen Arm erhielten konsequent Chemotherapie; überdies war die klinische Gesamt-Betreuung weitgehend festgelegt. Im Standard-Arm wurden die Patienten nach Gutdünken supportiv behandelt, mitunter arbiträr mit systemischer Chemotherapie. Die Resultate sprachen für einen Vorteil der regionalen Chemotherapie. Da die Betreuung der Patienten im Kontrollarm offensichtlich sehr viel weniger standardisiert und kontrolliert war, kann der klinische Vor-

teil im experimentellen Arm nicht mit Sicherheit auf die regionäre Chemotherapie an und für sich zurückgeführt werden. Leider wurde in dieser (in einer renommierten Zeitschrift publizierten) Studie ein neues klinisch sauber kontrolliertes Therapiekonzept mit einer Kraut-und-Rüben-Politik im Kontrollarm verglichen, und die Aussagekraft dadurch geschmälert.

Subgruppen-Analysen in randomisierten Studien. Obwohl die Einschlusskriterien randomisierter Studien die Selektion der Patienten vorschreiben, ist die behandelte Patientenpopulation schliesslich keineswegs homogen. Eine Studie beispielsweise, die sich mit der adjuvanten Therapie prämenopausaler Patientinnen mit nodal-positivem Mammakarzinom befasst, umfasst im Rahmen dieser Einschlusskriterien ein weites Spektrum an Patientinnen mit sehr unterschiedlichen prognostischen Faktoren. So ist die Prognose einer 35jährigen Patientin mit mehr als 4 befallenen Lymphknoten und negativen Hormonrezeptoren wesentlich schlechter als die Aussichten einer um 15 Jahre älteren Frau mit positiven Hormonrezeptoren und lediglich einer axillären Lymphknotenmetastase; obwohl beide die genannten Selektionskriterien erfüllen. Es liegt nahe, in derartigen Studien sekundär Subgruppen auszulesen und gesondert zu analysieren; beispielsweise Subgruppen, die sich bezüglich wesentlicher prognostischer Faktoren unterscheiden (Hormonrezeptor-Status, Lymphknotenstatus beim Mammakarzinom). Diese Analysen sind gängig, oftmals wertvoll, gelegentlich aber auch problematisch.

Die randomisierte ISIS-2-Studie verglich bei 17 187 Patienten mit Herzinfarkt die Behandlungsmöglichkeiten Streptokinase, Aspirin, Streptokinase und Aspirin, mit einem Kontrollarm ohne diese spezifischen Behandlungsoptionen [2]. Die Biostatistiker der Studie führten im Rahmen der Datenauswertung eine gezielte Subgruppen-Analyse durch. Sie wählten Kriterien, die klinisch interessant waren (Geschlecht, Diabetes in der Anamnese, Alter), aber auch Kriterien, die mit einiger Sicherheit für Patienten mit Herzinfarkt irrelevant sein dürften (Sternzeichen der Patienten gemäss Geburtsdatum) (Tab. 2). Offensichtlich schützt Aspirin nicht vor Tod durch Herzinfarkt, falls der Patient unter dem Sternzeichen *Waage* oder *Zwillinge* geboren wurde. Trotzdem würde wohl niemand bei dieser Patientensubgruppe auf Plättchenhemmer verzichten. Das Argument, Subgruppen-Analysen würden in der Regel mit klinisch relevanten Selektionskriterien durchgeführt, schützt nicht mit Sicherheit vor derartigen Fehlern.

Ergo: «Don't do subgroup analyses – if you do, don't believe the results.» Man mag sich dieser sehr strengen Sicht einzelner Biostatistiker nicht in jedem Fall anschliessen. Viele Daten aus Subgruppen-Analysen sind klinisch tatsächlich wertvoll. Die heute unbestrittene Tatsache beispielsweise, dass Patientinnen mit Mammakarzinom und negativen Hormonrezeptoren ein besonderes klinisches Problem darstellen, wurde ursprünglich mittels Subgruppenanalysen ermittelt. Die beste Lösung dieses Problems besteht darin, Daten aus Subgruppen-Analysen ihrerseits in prospektiver Weise zu überprüfen.

Statistische «power» einer randomisierten Studie. Dem Uneingeweihten bleiben statistische Überlegungen und Berechnungen von Studienresultaten oft ein Buch mit sieben Siegeln. Die Methodikabschnitte eines Artikels versichern uns beispielsweise, man hätte mit «Fisher's test» exakt nach Signifikanz gefischt. Hätte man vielleicht «Poisson» für diesen Zweck bemühen sollen? Oder wäre die Quadratur des Datenkreises mit dem «Chi square test» besser gelungen? Dem einfachen Kliniker ist es oft unmöglich, zu entscheiden, ob die gewählten Statistikttests «indiziert» waren oder nicht. Es ist für mathematisch-statistische Laien meist auch nicht überprüfbar, ob in einer Studie wirklich genügend Patienten eingeschlossen wurden, und ob Unterschiede in den Therapieresultaten rechnerisch niet- und nagelfest abgesichert werden konnten. Immerhin können ein paar Überlegungen nützlich sein, die statistisch naiven Lesern (zu denen sich die Autoren dieses Artikels zählen) zugänglich sind.

- Man überlege sich, ob eine neue Therapie *prima vista* das Potential für einen beträchtlichen oder eher für einen bescheidenen Fortschritt mit sich bringt. Je kleiner der zu erwartende Therapieeffekt, desto grösser ist die benötigte Zahl an Studienpatienten, um ihn zu belegen.
- Studien, die mit einer kleinen Patientenzahl einen beträchtlichen Fortschritt aufzeigen, können

- a) tatsächlich einen eigentlichen Durchbruch darstellen;
- b) statistisch schwach sein, so dass der Effekt überschätzt wird; oder
- c) falsch positiv sein.

Die Optionen b) und c) sind mindestens in der Onkologie (leider) häufiger als a.

- Legt die Publikation über alle eingeschlossenen Patienten Rechenschaft ab? In jeder Studie kann es vorkommen, dass einige Patienten eingeschlossen werden, welche die Einschlusskriterien nicht erfüllen. Ihre Zahl sollte möglichst gering sein (nicht mehr als einige % der Gesamtpatientenzahl).
- Sind alle Patienten auswertbar? Wieviele nicht? Wieviele Patienten erhielten die geplante Therapie, bei wievielen wurde sie vorzeitig abgebrochen? Wieviele Patienten konnten in den Nachkontrollen nicht mehr erfasst werden («lost to follow-up»)? Je höher die Zahl an Patienten, deren Daten nicht auswertbar oder nicht vollständig sind, desto problematischer sind die Resultate einer Studie.

Meta-Analysen

Meta-Analysen sind in die Mode gekommen. Sehen die einen in diesem Instrument einen entscheidenden Fortschritt für die Auswertung klinischer Studiendaten, so werden auch Stimmen laut, die diese Methodik als «lots of fancy statistics on lots of garbage data» bezeichnen. Meta-Analysen fassen Daten aus zahlreichen randomisierten Studien mit ähnlicher Fragestellung zusammen, mit dem Ziel, eine statistische «power» zu gewinnen, wie sie in einzelnen Studien nur sehr selten erreicht wird [3]. Damit sind sie speziell für die Absicherung quantitativ geringer, aber dennoch relevanter Therapieeffekte geeignet. So können z.B. alle Studien zusammengefasst werden, die sich mit der Frage befassen, ob eine adjuvante Chemotherapie beim nodalnegativen Mammakarzinom nützlich ist. Für eine derartige Meta-Analyse kandidieren Studien, die Chemotherapie mit einem Kontrollarm vergleichen; welche Art einer Chemotherapie in den einzelnen Studien eingesetzt wird, ist dagegen unerheblich. Damit dienen Meta-Analysen vor allem dem Zweck, therapeutische Prinzipien zu untersuchen, unter gezielter Missachtung von Einzelheiten, wie z.B. einzelnen Chemotherapie-Regimes. Der Wert einer Meta-Analyse steht und fällt mit der Auswahl von Studien (Abb. 1). Idealerweise sollte eine Meta-Analyse zu einem bestimmten Thema alle Studien einschliessen, unabhängig davon, ob sie publiziert sind oder nicht, und unabhängig von ihren Resultaten. Aus einer Publikation sollte die Strategie der Autoren, passende Artikel zu identifizieren, hervorgehen.

Tabelle 2. Subgruppen-Analyse in der ISIS-2-Studie.

Sternzeichen	Aspirin-Effekt auf die Mortalität zwischen 0–30 Tagen nach Herzinfarkt	
	Anzahl Todesfälle Aspirin-Gruppe vs. Placebo-Gruppe	Statistische Signifikanz
Waage oder Zwillinge	150 / 8587 vs. 147 / 8600	nicht signifikant
Alle anderen Sternzeichen	564 / 8587 vs. 869 / 8600	p < 0,000 000 1

Probleme der Publikation von Studien

Abstracts. «Abstracts», d.h. kurze Zusammenfassungen von Studiendaten sind der übliche Weg, um sich die Möglichkeit zur Datenpräsentation an Kongressen zu verschaffen. Oft werden Daten aus laufenden Studien oder «unreife» Frühdaten nach kurzer Beobachtungsdauer zur Kongress-Präsentation angemeldet. Die Frage stellt sich, wie oft Daten, die als Abstracts präsentiert und in Kongress-Bänden publiziert wurden, später in Form von formellen und sauber ausgearbeiteten Originalpublikationen in die Literatur eingehen. Weniger oft, als uns lieb ist!

Im Kongressband der American Society of Clinical Oncology (ASCO) aus dem Jahre 1984 wurden etwa 200 Abstracts ausgelesen [5]. Nur 60% dieser Abstracts wurden später in einer definitiven Fassung als Originalartikel publiziert. Drei Viertel aller Abstracts, die am Kongress präsentiert werden konnten, schafften es bis zur Publikation, während bloss die Hälfte der im Kongressband nur abgedruckten aber nicht präsentierten Ab-

stracts das Licht der Welt in Form eines vollen Artikels erblickten. Falls ein Abstract über positive Daten berichtete (zum Beispiel eine Therapieverbesserung mit einem neuen Medikament), betrug seine Chance für eine mündliche Kongress-Präsentation und spätere Publikation 60–75%; bei negativen Daten lediglich etwa 30%. Bei 80% aller Abstracts über klinische Studien wurden in der späteren Publikation dieselben Daten und Schlussfolgerungen dargestellt; bei einem Fünftel der Fälle ergaben sich gewichtige Divergenzen in den Daten und Aussagen zwischen Abstract und späterer Publikation.

Abstracts sollten mit der nötigen kritischen Distanz interpretiert werden. Sie bieten meist keine genügende Basis, um neue Behandlungsstandards festzulegen.

Publikationschancen für positive und negative Daten. Positive Daten klinischer Studien sind leichter publizierbar als negative Resultate [6]. Überdies ist die Motivation, bei Vorliegen positiver Daten den mitunter beschwerlichen Weg bis zur Publikation zu gehen, höher, als wenn negative Daten erarbeitet werden. Dieses an sich verständliche Prinzip birgt die Gefahr, dass klinisch relevante negative Resultate in der Literatur untergehen. Es wäre wichtig zu belegen, dass ein neues Medikament keine Vorteile gegenüber einer eingebürgerten und vertrauten, oft billigeren Therapie bietet. Dies wird besonders kritisch bei der Selektion von Studien für sogenannte Meta-Analysen (siehe oben).

Sprach-«Bias» in wissenschaftlichen Zeitschriften. Publizieren ist eine akademische Tätigkeit, die für eine brillante Karriere essentiell ist («publish or perish»). Den Publikationen in international anerkannten, streng selektierenden Journals mit sogenannten «peer-review»-System kommt dabei besonderes Gewicht zu. Artikel in lokalen Zeitschriften, in der Regel in der Landessprache verfasst, sind daneben weniger gewichtige Elemente eines Curriculum vitae oder einer Publikationsliste. Eine interessante Studie ist der Frage nachgegangen, inwiefern sich Arbeiten über klinische Studien aus der Feder deutschsprachiger Autoren unterscheiden, je nachdem ob sie international in Englisch oder lokal in Deutsch publiziert wurden [7].

Es wurden Studienleiter in der Schweiz, in Deutschland und in Österreich identifiziert, die klinische randomisierte Studien sowohl in Deutsch als auch in Englisch publizierten. Für Studien mit positiven Resultaten (z.B. Berichte, die eine statistisch signifikante Differenz zugunsten einer neuen Behandlung aufzeigen konnten) war die Chance zur Pu-

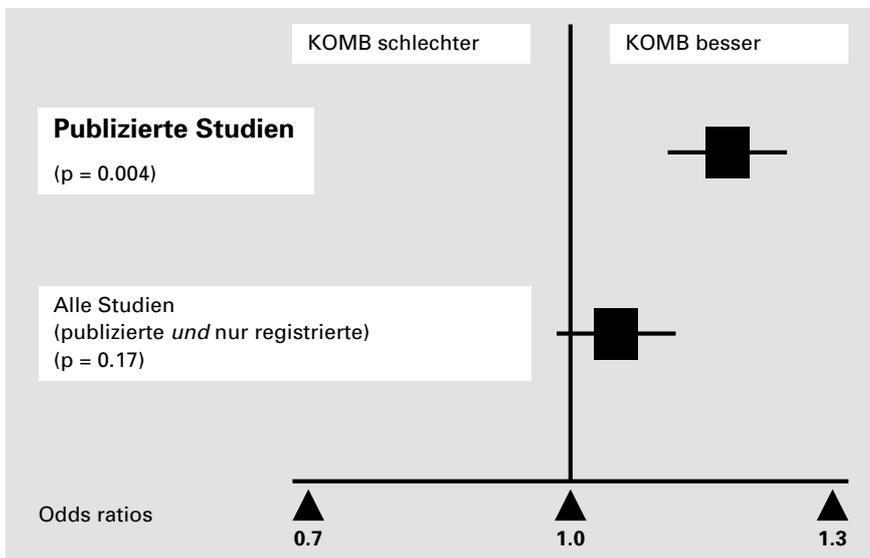


Abbildung 1. Meta-Analyse zur Frage, ob beim inoperablen Ovarialkarzinom Kombinations-Chemotherapien (KOMB) besser seien als Chemotherapien mit nur einem Zytostatikum. Eine «odds ratio» von 1,0 impliziert, dass beide Therapieoptionen gleichwertig wären. Die schwarzen Quadrate zeigen die Mittelwerte der Therapieeffekte aller berücksichtigten Studien an: Werte >1,0 implizieren einen Vorteil für die Variante Kombinationschemotherapie, Werte unter 1,0 einen Vorteil für Monotherapie. Die waagrechten Balken zeigen den Streubereich an [4]. Die Auswertung aller *publizierten* Studien zu diesem Thema spricht für einen Vorteil der Kombinations-Chemotherapien. Werden aber alle *registrierten* Studien berücksichtigt (publizierte und nicht-publizierte), so fällt das Resultat weit weniger eindeutig aus. Der hauptsächlichste Grund für diesen Effekt ist, dass sich unter den registrierten Studien einzelne Studien finden, die keinen Vorteil für eine Chemotherapie mit mehreren Zytostatika fanden, aber (deswegen?!) nicht publiziert wurden oder nicht publiziert werden konnten (siehe auch Abschnitt über «publication bias» von positiven und negativen Daten).

blikation in einer anerkannten englischsprachigen Zeitschrift 3–4 × höher als für negative Studien, die (wahrscheinlich nach Ablehnung im einen oder anderen internationalen Journal) schliesslich deutsch publiziert (teils «vergraben») wurden [7].

Negative Resultate mögen jedoch durchaus relevant sein; beispielsweise mag es wesentlich sein, wenn für eine neue (oft teurere) Therapie kein Nachweis einer besseren Wirkung erbracht werden kann. Das Risiko besteht, dass derartige Berichte in der englischsprachigen Literatur unterrepräsentiert sind. Überdies ist bei klinischen Studien aus verschiedenen statistischen Überlegungen (auf die hier nicht im Detail eingegangen wird) das Risiko falsch-positiver Resultate beträchtlich höher als das Risiko falsch negativer Daten. Die englischsprachige Literatur mit ihrer Vorliebe für die Publikation positiver klinischer Daten setzt sich somit einem höheren Risiko aus, falsche Resultate zu publizieren, als die weniger hoch eingeschätzten deutschsprachigen Zeitschriften, die negative «trials» zur Publikation akzeptieren.

Die besonderen Probleme der Übersichtsartikel (Reviews). Übersichtsartikel sind ebenso wenig vor Problemen gefeit wie Originalartikel. Der Leser erwartet in der Regel eine gewichtete Zusammenfassung der Literatur zum gewählten Thema und abgewogene, begründete Schlussfolgerungen. Übersichtsartikel können in «erzählerischem» Stil abgefasst sein (das vorliegende Werk ist ein Beispiel dafür), oder sie können als Meta-Analysen angelegt sein. Meta-Analysen sind von Vorteil, wenn therapeutische Fragen zur Debatte stehen, da die Methodik der Meta-Analysen weitgehend ausschliesst, dass ein Verfasser einer Review

einzelne Studienresultate arbiträr hervorhebt und andere herunterspielt oder gar nicht berücksichtigt. Sie gewährleisten, dass die Auswahl der einbezogenen Studien transparent ist, und sie garantieren eine nachvollziehbare Verarbeitung der Daten. In diesem Sinne sind sie Übersichtsartikeln vorzuziehen, die Daten randomisierter Studien im Erzählstil kommentieren [8].

Die Beurteilung der Qualität einer im Erzählstil verfassten Übersichtsarbeit («narrative review article») ist nicht einfach, und es ist erstaunlich, dass die Diskussion über dieses Problem erst vor wenigen Jahren in Schwung gekommen ist. Übersichtsarbeiten wie der vorliegende Text können nicht so ohne weiteres standardisiert werden wie die Zusammenfassung randomisierter Studien mit dem Instrument der Meta-Analyse. Das Potential für bewusste oder unbewusste Gewichtung der Daten in die eine oder andere Richtung («bias»), für arbiträre Interpretation und für gezielte Selektion der berücksichtigten Arbeiten (mitunter mit Unterdrückung von Arbeiten gegnerischer Gruppen) ist vorhanden. Immerhin können einige Erfordernisse an die Qualität eines Artikels formuliert werden. So sollten die Fragestellungen eingangs klar dargelegt werden, und die Technik der Literatursuche und -auslese sollte kurz angesprochen werden.

Schlussfolgerungen für den Alltag

Nicht alles, was gedruckt ist, stimmt! Sogar renommierte Zeitschriften mögen sich trotz strenger «peer-review» in der Wahl von Artikeln, die sie zur Publikation annehmen, irren. Es ist unmöglich, auch für den Fachmann in einem Gebiet, bei jeder Publikation mit Sicherheit herauszudestillieren, ob die berichteten Erkenntnisse richtig sind, und damit von nachhaltiger Wirkung für die Praxis, ob sie Eintagsfliegen darstellen, oder ob sie schlicht falsch sind. Immerhin können ein paar Merksätze für die Beurteilung von Artikeln über klinisch-therapeutische Studien zusammengestellt werden. Wir nehmen nicht für uns in Anspruch, dass unser Übersichtsartikel in Inhalt und Form über jeden Zweifel erhaben sei, und dass wir alle Fussangeln, die wir bei anderen Arbeiten gefunden haben, selbst elegant umgangen hätten. Wir implizieren andererseits in keiner Weise, dass fast alles, was gedruckt wird, grundsätzlich einmal verdächtig ist. Wir hoffen jedoch, dass wir – wenn auch lückenhaft – einige Denkanstösse zu einer kritischen Lesetechnik für medizinische Fachliteratur, am Beispiel onkologischer Studien und anderer Arbeiten, haben vermitteln können. Warum wenden Sie unsere Anregungen zur kritischen Durchsicht der Fachliteratur nicht übungshalber auf unsere Artikel an?

Quintessenz

- **Abstracts** und Kongress-Mitteilungen genügen nicht, um neue klinische Standards aufzustellen.
- Man beachte den «**level of evidence**» einer Studie. Je höher er ist, desto zuverlässiger sind die publizierten Daten.
- Man vergewissere sich, ob Studienresultate durch **unabhängige** Studien-Gruppen in einer separaten Patientenpopulation bestätigt werden konnten.
- Man achte auf die **Vergleichspopulation**, an der neue klinische Resultate gemessen werden. Entsprechen die publizierten Resultate des Standardarms, der herkömmlichen Therapie, den Erfahrungen und den Erwartungen?
- Man achte bei **Übersichtsartikeln** («reviews») darauf, ob die Methodik (z.B. Technik der Literatursuche), auf der die Aufarbeitung der diskutierten Daten beruht, im Artikel dargestellt wird.

Literatur

- 1 Allen-Mersh TG, Earlam S, Fordy C, Abrams K, Houghton J. Quality of life and survival with continuous hepatic-artery floxuridine infusion for colorectal liver metastases. *Lancet* 1994;344:1255-60.
- 2 ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;II:349-60.
- 3 Egger M, Davey Smith G, Andrew N Phillips. Meta-analysis: principles and procedures. *Br Med J* 1997;315: 1533-53.
- 4 Symes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11-29.
- 5 De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Ann Oncol* 1992;3:187-91.
- 6 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267: 374-8.
- 7 Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; 350:326-9.
- 8 Williams CJ. The pitfalls of narrative reviews in clinical medicine. *Ann Oncol* 1998;9:601-5.