pharma-kritik

AZ 9500 Wil

ISSN 1010-5409

Jahrgang 11 Nr. 14 28. Juli 1989

Übersicht

Klinische Studien kritisch durchleuchtet

Anleitung zum Lesen von Therapieberichten

E. Gysling

Manuskript durchgesehen von F. Follath, J. Osterwalder und J. Reichen

Die Lektüre von medizinischen Fachzeitschriften ist eines der wichtigsten Mittel ärztlicher Fortbildung. Die dafür verwendete Zeit kann gemäss einer eigenen Umfrage und nach publizierten Untersuchungen[1,2] auf durchschnittlich zwei Stunden pro Woche geschätzt werden. Dies sind immerhin rund zwei Arbeitswochen jährlich, die so nutzbringend wie möglich eingesetzt werden sollten.

Anderseits nimmt die Fülle medizinischen Fachwissens ständig so sehr zu, dass es immer schwieriger wird, den für unsere ärztliche Tätigkeit notwendigen «Wissenskern» auf dem aktuellen Stand zu halten. *Übersichtsartikel,* welche ein Teilgebiet auf knappem Raum ausleuchten, helfen uns, mögliche Defizite zu beheben. Es gibt aber gute Argumente, weshalb wir immer wieder *auch Originalarbeiten* über klinische Studien lesen sollten. Eine eigene Meinung zur Frage, ob wir eine neue Therapie (z.B. ein neues Medikament) *bei unseren Patienten* einführen sollen, können wir uns am besten anhand der Originalarbeiten bilden.

Wie ist es möglich, aus den vielen publizierten Studien die wichtigen herauszufinden und wie gelingt es uns als Nicht-Spezialisten, diese Berichte kritisch zu werten? Zu diesen Fragen sind in den letzten Jahren mehrere Leitfaden[3-7] erschienen. In der folgenden kurzen Anleitung zum Lesen von Therapieberichten stütze ich mich weitgehend auf diese Arbeiten. Grundbegriffe wie Mittelwert, Median und Standardabweichung und entsprechende statistische Tests werden nicht erläutert; im Vordergrund steht die Beurteilung der *praktisch-medizinischen Relevanz* der Studien. Es ist klar, dass eine genauere biostatistische und epidemiologische Analyse entsprechende Spezialkenntnisse erfordert, die hier nicht vermittelt werden können.

Für und wider Originalarbeiten

Die in Tabelle 1 zusammengestellten Argumente für und wider das Lesen von Originalarbeiten illustrieren unsere Befürchtungen, ohne enormen Zeitaufwand niemals «auf dem Laufenden» bleiben zu können. (Tatsächlich gibt es Studien, die eine Tendenz der Ärzte aufzeigen, bei dem in den Studenten- und Assistentenjahren Gelernten zu verharren.[8]) Nur wenn es uns gelingt, mittels individueller Selektion die richtige Auswahl zu treffen, wird die zum Lesen verwendete Zeit auch unseren Patienten zugute kommen.

Die Zeitschriften

Es lohnt sich, die Zeitschrift(en), die wir regelmässig lesen wollen, sorgfältig auszuwählen. Um innerhalb der kurzen zur Verfügung stehenden Zeit auch gelegentlich eine Originalarbeit über eine klinische Studie lesen zu können, müssen unsere Zeitschriften hohen Qualitätsanforderungen genügen. Gute Zeitschriften enthalten *echte* Originalarbeiten und Übersichten mit einem hohen Anteil an

Tabelle 1: *Für und wider das Lesen von Originalarbeiten (nach [3])*

Vorteile
Aktualität der Information
Kritische redaktionelle Durchsicht der Studien
Dokumentation der Zuverlässigkeit der Daten
Konsistenz der Präsentation (Methodik, Resultate)
Nachteile
Zu grosse zeitliche Beanspruchung
Zu stark limitierte Fragestellung einzelner Studien
Abonnementskosten bzw. ungenügende Verfügbarkeit der Zeitschriften
Sprachliche Probleme (fremdsprachige Zeitschriften)

praxisrelevanten Beiträgen. Diese Beiträge sollen nicht zu lang, gut dargestellt und leicht lesbar sein. Zeitschriften, welche ihre Texte einer «Peer Review» unterwerfen, bieten eine gewisse Gewähr dafür, dass nur kritisch gesichtete Studien veröffentlicht werden.

Das Prestige der «grossen» medizinischen Zeitschriften *in englischer Sprache* bewirkt, dass diese Zeitschriften aus den besten Arbeiten auswählen können. Es hat keinen Sinn, diese Tatsache zu beklagen; sie lässt sich kaum ändern. Über neue Entwicklungen wird oft zuerst in einer der grossen «allgemeinmedizinischen» Zeitschriften aus England oder aus den USA berichtet. Kolleginnen und Kollegen, die gern in englischer Sprache lesen, kann deshalb empfohlen werden, eine dieser Zeitschriften zu abonnieren.

Die zahlreichen sogenannten Fachzeitschriften, die uns *gratis* ins Haus geschickt werden, enthalten praktisch nie Originalarbeiten in akzeptabler Qualität. In den meisten Fällen ist die für diese Blätter aufgewendete Zeit verlorene Zeit.

Ein Problem betrifft alle, auch die «grossen» Zeitschriften: es werden viel *zu selten Studien mit negativen Resultaten* veröffentlicht. Klinische Studien, die einen Fortschritt zu bringen scheinen, sind offensichtlich viel attraktiver. Wenn wir daher über eine erfolgversprechende neue Therapie lesen, sollten wir immer daran denken, dass diese Therapie vielleicht in vielen anderen (unveröffentlichten) Studien negative Resultate ergeben hat. In diesem Zusammenhang sind die Editorials, wie sie besonders von den englischsprachigen Zeitschriften gepflegt werden, wertvoll. Sie helfen uns oft, eine möglicherweise zu optimistische Schlussfolgerung einer Studie in der richtigen Perspektive zu sehen.

Die wichtigsten Auswahlkriterien

Eine Handvoll von Auswahlkriterien genügt, die Zahl lesenswerter Artikel auf ein Mass zu reduzieren, welches sich bewältigen lässt. Diese Kriterien beziehen sich in erster Linie auf Arzneimittel-Studien, können aber auch auf andere Therapiestudien, zum Teil auch auf klinische Studien anderer Natur angewendet werden. (Schwieriger zu beurteilen sind sogenannte Meta-Analysen, in denen kleinere – für sich allein wenig aussagekräftige – Studien zusammengefasst sind. Gute Meta-Analysen können aber durchaus dazu beitragen, komplexe therapeutische Fragen besser zu beantworten.)

Wichtig ist die Erkenntnis, dass *die Aussagekraft einer Studie* ganz entscheidend *von der Studienmethodik abhängt.* Bei inadäquatem Studiendesign sind auch die schönsten «statistisch signifikanten» Resultate wertlos.

Um darüber zu entscheiden, ob die folgenden Fragen bejaht werden können, orientiert man sich zunächst am Titel und an der Zusammenfassung einer Arbeit. Dies erlaubt, mit minimalem Zeitaufwand die Studien zu eliminieren, welche nicht lesenswert sind. Bei denjenigen Studien aber, welche diesen «Kurztest» bestehen, muss die Beschreibung der Methodik genauer angesehen werden.

Handelt es sich um eine kontrollierte Studie?

Nur der Vergleich zwischen verschiedenen Behandlungen kann zeigen, welche die beste ist. Dieser Vergleich ist aber nur dann aussagekräftig, wenn die Zuteilung der Patienten zu einer Behandlung *randomisiert* – nur durch den Zufall bestimmt – ist. Alle anderen Verfahren (z.B. die alternierende Zuteilung) führen zu einer Verfälschung der Resultate. Im Idealfall soll die Zuteilung ohne das Wissen von Patient und Arzt, also *doppelblind* erfolgen. Verschiedene Tricks ermöglichen Doppelblindstudien auch in Fällen, die für dieses Design a priori ungeeignet erscheinen. Besonders bei nicht-medikamentösen Verfahren ist aber kaum eine Placebobehandlung zu realisieren; auch Medikamente mit charakteristischen Nebenwirkungen entziehen sich unter Umständen einer blinden Austestung. In Tabelle 2 sind Grundregeln zusammengestellt, welche erlauben, die Aussagekraft von Studien einzuschätzen.

Tabelle 2: *Aussagekraft von klinischen Studien (nach [9])*

A	*Sehr hohe Aussagekraft* Grosse randomisierte Studien mit eindeutigen Resultaten
B	*Mässige Aussagekraft* Kleine randomisierte Studien mit unsicheren Resultaten (und entsprechend erhöhten Fehlerrisiken)
C	*Geringe Aussagekraft* Nicht-randomisierte Studien mit gleichzeitigen oder historischen Kontrollgruppen. Fallberichte ohne Kontrollen.

Ist die Fragestellung der Studie für meine Patienten wichtig?

Diese Frage ist von Bedeutung, um eine individuelle Selektion treffen zu können. Hauptamtlich im Spital tätige Ärzte werden andere Artikel wählen als Allgemeinpraktiker; eine Ophthalmologin benötigt andere Information als eine Kinderpsychiaterin. Oft sind Studien von wenig Belang für die Praxis, weil ihre *Dauer* realitätsfremd ist. Bei Medikamenten, die üblicherweise langfristig verabreicht werden müssen, vermitteln Studien mit Einzeldosen oder kurzfristiger Anwendung keine adäquate Information. Es gibt auch eine beträchtliche Zahl von Studien, die höchstens für die hinter der Studie stehende Firma von Bedeutung sind: ich denke an Studien mit «neuen» Medikamenten, welche sich von den bereits vorhandenen nicht nennenswert unterscheiden und die lediglich aus kommerziellen Gründen eingeführt werden.

Ist die Studienpopulation genügend gross?

Eine Studie muss nicht viele Patienten umfassen, wenn der Unterschied zwischen den verglichenen Behandlungen

gross ist. Meistens ist die Wirkung einer neuen Behandlung aber nicht so dramatisch und ihre Vorteile können nur in verhältnismässig grossen Studien demonstriert werden. Es ist deshalb wichtig, dass *vor* Studienbeginn berechnet wird, wieviele Probanden in die Studie einbezogen werden müssen. Damit kann gesichert werden, dass sich wahre Unterschiede zwischen den Vergleichsgruppen bzw. das Fehlen eines Unterschiedes durch einen statistischen Test adäquat prüfen lassen.
Ein sogenannter *Typ II-Irrtum* liegt vor, wenn sich infolge einer kleinen Probandenzahl in einer Studie kein statistisch signifikanter Unterschied errechnen lässt. Viele kleine Studien leiden an diesem Übel; da sich kein statistisch signifikanter Unterschied zwischen den verglichenen Behandlungen ergeben hat, wird fälschlicherweise geschlossen, die Behandlungen seien gleichwertig.

Ist das Resultat klinisch relevant?

Ein statistisch signifikanter Unterschied ist nicht notwendigerweise mit einem klinisch oder praktisch relevanten Unterschied identisch. In den üblicherweise angegebenen Resultaten (z.B. $p < 0,05$) kommt nicht zum Ausdruck, wie gross die «Genauigkeit» und biologische Wertigkeit der Resultate ist. Diese hängt unter anderem davon ab, wieviele Probanden untersucht wurden und wie gross die Variabilität der untersuchten Grössen ist.
Eine bessere Information vermittelt das *Vertrauensintervall.* Diese Grösse entspricht dem Bereich, in welchem sich der wahre Wert mit einer bestimmten Wahrscheinlichkeit (z.B. 95%) findet; es kann aus den Einzelwerten und der Zahl der Probanden errechnet werden.[10,11] Das Vertrauensintervall vermittelt eine medizinisch «einleuchtende» Dimension. Ein Beispiel: In einer Studie senkt ein neues Antihypertensivum den diastolischen Blutdruck durchschnittlich um 6 mm Hg mehr als die Vergleichssubstanz (statistisch signifikant). Das Vertrauensintervall, in welchem sich der wahre Wert mit 95% Wahrscheinlichkeit befindet, liegt zwischen 1,1 und 10,9 mm Hg. So ist es uns möglich, uns ein eigenes Bild von der praktischen Bedeutung der gefundenen Differenz zu machen.

Weitere Beurteilungskriterien

Die bisher genannten Kriterien dienen dazu, «die Spreu vom Weizen» zu trennen. Hat man sich einmal zum Lesen einer Arbeit entschlossen, gilt es herauszufinden, ob die Studie einer kritischen Analyse standhält. Dabei muss in erster Linie die Anwendbarkeit der Resultate auf die medizinische Realität in Praxis und Klinik geprüft werden.

Welche Patienten wurden in die Studie aufgenommen?

Die Definition der Patienten, welche in die Studie aufgenommen *wurden,* sollte aus dem Bericht klar hervorgehen. Alter und Geschlecht der Patienten sind wichtig, aber auch, ob es sich um Spital- oder ambulante Patienten handelte. Von besonderer Bedeutung sind die *Ausschlusskriterien.* Schweregrad der Erkrankung und eventuelle Begleiterkrankungen müssen ebenfalls eindeutig umrissen sein, um uns zu erlauben, die «Studienpatienten» mit unseren eigenen zu vergleichen. Studien, bei denen komplexe «Diagnoseskalen» zur Anwendung gelangen (z.B. bei Depressionen) sind schwierig zu interpretieren. Es ist auch daran zu denken, dass der Spezialist oft eine andere Auswahl von Patienten behandelt als der Allgemeinmediziner oder -internist. Was für eine ausgewählte Gruppe gilt, hat nicht immer allgemeine Gültigkeit.

Ist die zum Vergleich dienende Behandlung gut gewählt?

Es ist richtig, ein neues Medikament mit einem *Placebo* zu vergleichen, wenn die Wirksamkeit des neuen Mittels noch in Frage steht oder wenn keine andere wirksame Behandlung bekannt ist. Eine Placebokontrolle ist auch dann sinnvoll, wenn zwei aktive Substanzen bei einer Erkrankung verglichen werden, die oft spontan bessert. Verhältnismässig wenige Probanden genügen, um die Überlegenheit eines (aktiven) Arzneimittels gegenüber einer pharmakologisch inaktiven Substanz zu demonstrieren. Von viel grösserer Bedeutung ist aber der Vergleich mit einem aktiven Standard und zwar *mit der besten bisher bekannten Behandlung.* Substanzen, die gegenüber der bisherigen Behandlung keinen Fortschritt bringen, sind grundsätzlich ohne Interesse.

Wie wurden die Behandlungen durchgeführt?

Einzelheiten der Behandlungs-Modalitäten (Dosis, Relation der Verabreichung zu den Mahlzeiten, Möglichkeiten einer Zusatzmedikation bzw. einer Dosiserhöhung usw.) dürfen nicht vernachlässigt werden. Heikel ist z.B. die Frage, ob die *Dosis* des Vergleichsmedikamentes richtig festgelegt wurde. Sowohl eine relative Unterdosierung als auch eine Überdosierung können die Resultate verfälschen; praktisch ist es allerdings manchmal sehr schwierig, «gleichwertige» Dosen von zwei ähnlich wirkenden Medikamenten zu bestimmen. Im übrigen ist es wesentlich, sich auch über die *Compliance* Rechenschaft abzulegen, beispielsweise, indem die nicht verwendeten Tabletten gezählt werden.

Wie wurden die Wirkungen der Behandlungen beurteilt?

Die Beurteilung des Therapieerfolgs richtet sich nach der Fragestellung der Studie. In gewissen Fällen sind nur «harte» Daten (Tod, Infarktrezidiv usw.) aussagekräftig genug; in anderen Studien können oder müssen wir uns mit «weichen» Daten (z.B. Schmerzbeurteilung mittels visueller Analogskalen) begnügen. Wenn es sich um eine *Multizenter-Studie* handelt, so ist eine minutiöse Standardisierung unerlässlich. Soll die Studie eine Aussage zur Wirkungsdauer eines Medikamentes erlauben, so muss auch der *Zeitpunkt* von Messungen (z.B. einer Blutdruckmessung) genau festgelegt werden. Schliesslich ist darauf zu achten, dass nicht nur die erwünschten, sondern auch *unerwünschte Wirkungen* genau erfasst werden. Dieser letzte Punkt wird in kleineren Medikamentenstudien oft vernachlässigt. Die Art und Zahl unerwünschter Wirkungen unter der geprüften Behandlung muss mit den entsprechenden Resultaten in der Vergleichsgruppe (z.B. unter Placebo) verglichen werden.

Umfassen die Resultate alle Patienten, die in die Studie aufgenommen wurden?

Wenden wir uns nun den Resultaten zu, so gilt unser erstes Augenmerk der Vollständigkeit der Daten. Wenn uns Angaben über diejenigen Patienten fehlen, welche die Studie *nicht* regulär beendet haben, so gewinnen wir ein falsches Bild von der geprüften Behandlung. Oft sind es ja unerwünschte Wirkungen, die dafür verantwortlich sind, dass Patienten vorzeitig aus einer Studie ausscheiden. Je nach Studienendpunkt kann es wünschenswert sein, nicht nur die Gründe des Ausscheidens zu kennen, sondern bei der statistischen Analyse auch die ausgeschiedenen Patienten einzuschliessen. Wenn eine solche Analyse «by intention to treat» durchgeführt wird, erhält man naturgemäss ein weniger optimistisches, jedoch realitätsnäheres Resultat.

Enthält der Bericht alle wichtigen Einzelresultate?

Wenn wir Einsicht in die unveränderten, «rohen» Daten erhalten, ist es uns möglich, die Daten auch selbst zu interpretieren. Bei kleineren Studien sollten wir die Resultate jedes einzelnen Probanden sehen können; ist die Studie grösser, so lässt sich die Verteilung der Einzelresultate graphisch darstellen. Resultate, die nur Prozentzahlen (oder noch schlimmer, nur die Angabe «signifikanter» bzw. «nicht-signifikanter Unterschied») umfassen, sind ungenügend. Nicht selten werden der Metabolismus und damit die Wirkungsintensität eines Medikamentes von genetischen Faktoren modifiziert. Solche pharmakogenetischen Varianten können bewirken, dass in einer Studie einzelne Patienten viel mehr oder viel weniger auf ein Medikament reagieren.

Entspricht die Darstellung der Resultate ihrer praktischen Bedeutung?

Grundsätzlich sollten die Resultate ausschliesslich entsprechend den initial definierten Studienzielen analysiert werden. Wenn sich jedoch die einer Studie zugrundeliegende Hypothese nicht bestätigen lässt, sind die Studienautoren möglicherweise versucht, durch «geeignete» Unterteilung der Studienpopulation wenigstens in einer Untergruppe einen Nutzen zu demonstrieren. Auch in den letzten Jahren sind wiederholt Studien publiziert worden, bei denen nur dank einem «positiven» Teilresultat die Hypothese eines Behandlungsnutzens teilweise aufrechterhalten werden konnte. In vielen Fällen ist es aber beim Lesen eines Berichtes kaum möglich, festzustellen, ob bestimmte Analysen wirklich schon ursprünglich vorgesehen waren.

Überhaupt ist zu berücksichtigen, dass die Autoren einer Studie ein verständliches Bedürfnis haben, die Resultate in einem günstigen Licht erscheinen zu lassen. So wurde z.B. bei Lipidsenker-Studien besonders Wert auf die *relative* Senkung der Infarkthäufigkeit (z.B. um 18%) gelegt, während die *absoluten* Zahlen (eine Senkung von 9,8 auf 8,1%) doch viel bescheidener anmuten. Auch bei Graphiken gilt es aufzupassen, ob sie wirklich ein unverzerrtes Bild der gefundenen Daten wiedergeben.

Entsprechen die Schlussfolgerungen tatsächlich den Resultaten?

Auch bei den Schlussfolgerungen muss damit gerechnet werden, dass die Autoren einer Studie im allgemeinen zu einer relativ optimistischen Deutung der Resultate gelangen. Oft teilen andere Kliniker diese Beurteilung nicht uneingeschränkt. Vereinzelt kommt es auch vor, dass offensichtlich negative Resultate «positiv» umgedeutet werden. Je mehr Endpunkte in einer Studie statistisch analysiert worden sind, desto grösser ist das Risiko, dass sich dabei auch falsch-positive «Signifikanzen» ergeben. Problematisch sind auch Verallgemeinerungen, die zwar plausibel erscheinen, jedoch nicht von entsprechenden Daten gestützt sind. (Dies trifft z.B. auf viele Aussagen zu, die im Zusammenhang mit der medikamentösen Cholesterinsenkung gemacht werden.) Zu einer ausgewogenen Bilanz gehört es auch, dass relevante Nebenwirkungen in der Wertung berücksichtigt werden.

Nur wenn uns eine neue Behandlung nach kritischer Durchsicht aller vorliegenden Daten so sehr überzeugt, dass wir wünschten, gegebenenfalls auch damit behandelt zu werden, sollten wir sie an unseren Patienten anwenden.

Literatur

1 L. Curry und R.W. Putnam: Can. Med. Assoc. J.*124:* 563, 1981
2 Council on Medical Education: Assoc. Am. Med. Coll. Continuing Med. Educ. Newsl.*10:* 2, 1981
3 R.B. Hayes et al.: Ann. Int. Med.*105:* 149, 1986
4 Drug Ther. Bull.*23:* 1, 1985
5 Drug Ther. Bull.*23:* 5, 1985
6 Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre: Can. Med. Assoc. J.*124:* 555, 1981
7 Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre: Can. Med. Assoc. J.*124:* 1156, 1981
8 C.E. Evans et al.: JAMA*255:* 501, 1986
9 D.L. Sackett: Chest*95:* 2S, 1989
10 M.J.S. Langman: Br. Med. J.*292:* 716, 1986
11 M.J. Gardner und D.G. Altman: Br. Med. J.*292:* 746, 1986

Mitarbeiter dieser Ausgabe:

Prof. Dr. F. Follath, Medizinische Universitäts-Klinik B, Kantonsspital, CH-4031 Basel

Dr. J. Osterwalder, Zentrale Notfallstation, Kantonsspital, CH-9007 St.Gallen

Prof. Dr. J. Reichen, Institut für Klinische Pharmakologie der Universität, Murtenstr. 35, CH-3010 Bern

pharma-kritik

Herausgegeben von Etzel Gysling (Wil)
unter Mitarbeit von Renato Galeazzi (St. Gallen) & Urs A. Meyer (Basel)
Redaktion: Anne-Catherine Guex, Ulf Käsemodel, Urspeter Masche
Redaktionelle Mitarbeiter: B. Holzer (Thun), M.M. Kochen (München)
Verlagsmitarbeiter: Susanne Brändle-Schibenegg, Remo De Toffol
pharma-kritik erscheint zweimal monatlich
Bezugspreise: Jahresabonnement Fr. 78.- (Studenten Fr. 39.-),
Zweijahresabonnement Fr. 136.-, Einzelnummer Fr. 6.-
Infomed-Verlags-AG, Bergliweg 17, 9500 Wil, Telefon (073) 22 18 18
Druck: R.-P. Zehnder AG, Wil SG