

ÜBERSICHTSARBEIT

Kritisches Lesen wissenschaftlicher Artikel

Teil 1 der Serie zur Bewertung wissenschaftlicher Publikationen

Jean-Baptist du Prel, Bernd Röhrig, Maria Blettner

ZUSAMMENFASSUNG

Einleitung: Im Zeitalter der evidenzbasierten Medizin gehört das kritische Lesen wissenschaftlicher Artikel zu den grundlegenden Aufgaben eines Arztes. Dies ist erforderlich, um auf dem aktuellen Stand des Wissens zu bleiben und eine optimale Versorgung der Patienten sicherzustellen. Der vorliegende Artikel verfolgt das Ziel, eine einfache Einführung in das kritische Lesen wissenschaftlicher Artikel zu geben.

Methoden: Basierend auf einer Auswahl internationaler Literatur wird der Leser in das kritische Lesen von wissenschaftlichen Artikeln eingeführt. Zugunsten der Verständlichkeit wird auf die detaillierte Beschreibung statistischer Methoden verzichtet.

Ergebnisse: Es werden allgemeingültige Prinzipien zur Beurteilung von wissenschaftlichen Artikeln in der Medizin dargelegt. Grundlegende Kenntnisse zum Aufbau wissenschaftlicher Artikel, zum Studiendesign, zur statistischen Methodik, zu möglichen Fehlerquellen und Grenzen der Aussagekraft werden vermittelt. Auf Besonderheiten unterschiedlicher Forschungsbereiche wie Epidemiologie, klinische Forschung und Grundlagenforschung wird eingegangen, soweit das für das kritische Lesen wissenschaftlicher Artikel erforderlich ist. Zur Vertiefung in das Themengebiet wird auf weiterführende Literatur verwiesen.

Schlussfolgerung: Grundlegende methodische Kenntnisse sind zum Verständnis wissenschaftlicher Artikel notwendig.

Dtsch Arztebl Int 2009; 106(7): 100–5
DOI: 10.3238/arztebl.2009.0100

Schlüsselwörter: Publikation, kritisches Lesen, Entscheidungsfindung, Qualitätssicherung, Studie

Viele Ärzte haben heute trotz einer zunehmenden Publikationsflut immer weniger Zeit zum Lesen wissenschaftlicher Veröffentlichungen. Die Auswahl, Lektüre und kritische Beurteilung von Publikationen ist aber notwendig, um auf dem aktuellen Kenntnisstand zu bleiben. Diese Forderung ergibt sich auch aus den Grundsätzen der evidenzbasierten Medizin (1, 2).

Neben medizinisch-inhaltlichen Aspekten sind zum Verständnis und zur Bewertung auch Kenntnisse der statistischen Methodik notwendig. Leider werden selbst in der Wissenschaft nicht alle Begriffe immer richtig verwendet. So hat der Begriff „Signifikanz“ eine Inflation durch zu häufigen Gebrauch erfahren, weil signifikante (beziehungsweise positive) Ergebnisse leichter publiziert werden (3, 4).

Ziel dieses Artikels ist es, wesentliche Prinzipien der Beurteilung wissenschaftlicher Veröffentlichungen darzustellen. Diese Prinzipien gelten mit Ausnahme einiger studienspezifischer Besonderheiten gleichermaßen für experimentelle, klinische und epidemiologische Studien. Anhand der Literaturangaben ist eine Vertiefung in die Thematik möglich.

Entscheidungsfindung

Vor dem Lesen eines wissenschaftlichen Artikels muss die eigene Intention klar sein. Zur schnellen Information über den aktuellen Stand zu einem bestimmten Thema empfiehlt sich das Lesen eines aktuellen Reviews, zum Beispiel eines (einfachen) Übersichtsartikels, eines systematischen Reviews oder einer Metaanalyse.

Die in Übersichtsartikeln zitierte Literatur hilft, sich weiter mit der Fragestellung zu beschäftigen. Falls keine aktuellen, themenspezifischen Reviews vorliegen, ist zur Auffindung relevanter Fachartikel eine Recherche in einer Literaturdatenbank, zum Beispiel in PubMed, nötig.

Das regelmäßige Lesen von Fachzeitschriften bietet sich an, um auf dem aktuellen Kenntnisstand zu bleiben. Zur Entscheidungsfindung, welche Artikel lesenswert sind, tragen Überschrift und Zusammenfassungen bei. Die Überschrift vermittelt in knapper, präziser Form einen ersten Eindruck darüber, was den Leser im Folgenden erwartet. Die Zusammenfassung folgt dem gleichen Aufbau wie der Artikel und gibt die wesentlichen Aspekte der Publikation sehr verkürzt wieder. Das Lesen der Zusammenfassung ersetzt nicht die kritische Lektüre des gesamten Artikels, zeigt jedoch, ob es dem Autor gelingt, Fragestellung, Vorgehensweise, Ergebnisse und Schlussfolgerungen auf den Punkt zu bringen.

Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Johannes Gutenberg-Universität, Mainz: Dr. med. du Prel, MPH, Dr. rer. nat. Röhrig, Prof. Dr. rer. nat. Blettner

Struktur wissenschaftlicher Veröffentlichungen

Der Aufbau wissenschaftlicher Artikel ist prinzipiell immer gleich. Der Überschrift, der Zusammenfassung und den Schlüsselwörtern folgen die verschiedenen Abschnitte der eigentlichen Publikation. Diese gliedert sich in Einleitung, Methodenteil, Ergebnisteil, Diskussion (EMED-Format) sowie gegebenenfalls Schlussfolgerung und Literaturangaben. Inhalt und Sinn der einzelnen Abschnitte werden im Folgenden näher beschrieben.

Einleitung

Die Einleitung soll den Leser zu dem Gegenstand der Untersuchung hinführen. In diesem Abschnitt soll der aktuelle Kenntnisstand unter Zitieren neuerer Literatur zum Thema vermittelt und die Notwendigkeit zur Klärung der Fragestellung ersichtlich werden. Ergebnisse von zitierten Studien sollten detailliert mit Angaben von Zahlenwerten dargestellt werden. Ungünstig ist die Verwendung allgemeiner Floskeln, wie zum Beispiel „inkonsistente Ergebnisse“, „etwas besser“, und so weiter. Insgesamt sollte man den Eindruck gewinnen, dass der Autor die zitierten Artikel gelesen hat. Im Zweifelsfall bietet es sich an, diese selbst nachzulesen. In einer guten Veröffentlichung werden entscheidende Aussagen durch Literaturangaben belegt.

Idealerweise führt die Vorgehensweise in diesem Abschnitt vom Allgemeinen zu Speziellen. Bereits in der Einleitung wird ersichtlich, welche Fragestellung mit der Studie geklärt werden soll und warum das gewählte Studiendesign dafür geeignet ist.

Methoden

Der Methodenteil ist wichtig und ähnelt in gewisser Weise einem Kochbuch. Anhand der methodischen Aspekte sollen dem Leser gleich Rezepturen die notwendigen Kenntnisse vermittelt werden, die Studie zu wiederholen. Hier stehen die wesentlichen Informationen, die eine Bewertung der Validität der Studie erlauben (6). Der Methodenteil kann sich in Unterpunkte mit eigenen Überschriften gliedern, so können beispielsweise Labormethoden getrennt von statistischen Methoden beschrieben werden.

Alle Schritte der Planung, der verwendeten Stichprobe (zum Beispiel Patienten, Tiere, Zelllinien), der Durchführung sowie der statistischen Auswertung sollen im Methodenteil beschrieben werden: Wurde vor Beginn der Studie ein Studienprotokoll erstellt? Ging der eigentlichen Untersuchung eine Pilotstudie voraus? Werden Studienort und -zeitraum genannt? Hier sollte man auch erwähnen, dass die Studie mit Zustimmung der zuständigen Ethikkommission durchgeführt wurde. Wichtigstes Element einer wissenschaftlichen Untersuchung ist das Studiendesign. Wenn das Studiendesign aus irgendwelchen Gründen nicht akzeptabel ist, trifft dies auch auf den Artikel zu, unabhängig davon, wie die Daten analysiert wurden (7).

Die Wahl des Studiendesigns sollte begründet und verständlich beschrieben werden. Fehlt die Beschreibung wichtiger methodischer Aspekte, ist Misstrauen angebracht. Wird beispielsweise das Randomisierungsverfahren nicht genannt, was häufiger der Fall ist (8), kann nicht ohne weiteres davon ausgegangen werden, dass auch

KASTEN 1

Fragen zur Beurteilung der Methodik einer Studie

- Kann auf Grundlage des Studiendesigns die forschungsleitende Fragestellung beantwortet werden?
- Wird beschrieben, ob es sich um eine konfirmatorische (beweisführende), eine explorative (Hypothesen generierende) oder eine deskriptive (beschreibende) Untersuchung handelt?
- Welcher Studientyp wurde gewählt und ist dieser angemessen zur Beantwortung der forschungsleitenden Fragestellung?
- Wird der Endpunkt der Untersuchung genau definiert?
- Welche statistische Maßzahl wird zur Charakterisierung des Endpunktes verwendet? Wird, etwa bei epidemiologischen Studien, die Inzidenzrate (Neuerkrankungsrate), die Prävalenz (Krankenstand), die Mortalitätsrate (Anteil der Bevölkerung, der an einer Krankheit stirbt), die Letalität (Anteil der Erkrankten, der an einer Krankheit stirbt) oder die Hospitalisierungsrate (Anteil der Bevölkerung, der krankheitsbedingt ins Krankenhaus eingewiesen wird) angegeben?
- Wird das Untersuchungsgebiet, die Population und der Erhebungszeitraum (einschließlich Follow-up-Zeit) beziehungsweise das Zeitintervall zwischen den Untersuchungen detailliert beschrieben?

tatsächlich randomisiert wurde (7). Die statistische Methodik sollte hier nachvollziehbar dargelegt und komplexere statistische Kenngrößen und Verfahren verständlich beschrieben werden, ergänzt durch Hinweise auf Fachliteratur. *Kasten 1* beinhaltet weitere Fragen zur Bewertung des Methodenteils.

Eine Darstellung zu Studiendesigns und deren Einsatz ist bei Altman (7), Trampisch und Windeler (9) und bei Klug et al. (10) zu finden. Bei experimentellen Studien ist die genaue Beschreibung der Versuchsanordnung und -durchführung wichtig. Die Genauigkeit einer Methode, also Reliabilität (Präzision) und Validität (Richtigkeit), müssen angegeben werden. Bei klinischen Studien verbessert sich die Aussagekraft durch den Einbezug einer Kontrollgruppe (aktive Kontrolle, historische Kontrolle oder Placebokontrolle) und der randomisierten, das heißt zufälligen Zuteilung der Patienten auf die Studienarme. Durch eine Verblindung, durch die Behandlungs- und Beobachtungsgleichheit gewährleistet wird, kann die Qualität gleichfalls erhöht werden. Eine klinische Studie sollte in aller Regel eine Abschätzung der benötigten Patientenzahl vor Studienbeginn (Fallzahlplanung) beinhalten. Nähere Details zu klinischen Studien sind zum Beispiel im Buch von Schumacher und Schulgen (11) zu finden. Internationale Empfehlungen, speziell für die Berichterstattung in randomisierten, kontrollierten klinischen Studien, findet man in der aktuellen Fassung des CONSORT-Statement („consolidated standards of reporting trials“) (12).

Epidemiologische Studien lassen sich in Interventionsstudien, Kohortenstudien, Fall-Kontroll-Studien, Querschnittsstudien und ökologische Studien einteilen. *Tabelle 1* gibt einen Überblick darüber, wann welcher Studientyp am besten geeignet ist (13). Ein Merkmal guter Veröffentlichungen ist die genaue Beschreibung der Ein- und Ausschlusskriterien für die Studienpopulation. Wie hoch

TABELLE 1

Angemessener Studientyp für epidemiologische Untersuchungen (nach 13)

Ziel der Untersuchung	Studientyp
Untersuchung seltener Krankheiten wie Tumorerkrankungen	Fall-Kontroll-Studien
Untersuchung seltener Expositionen wie industrielle Chemikalien	Kohortenstudie in einer Bevölkerungsgruppe, in der die Exposition vorhanden ist
Untersuchung multipler Expositionen, wie etwa der gemeinsame Effekt von oralen Kontrazeptiva und Rauchen	Fall-Kontroll-Studien
Untersuchung multipler Endpunkte, wie das Sterberisiko aufgrund unterschiedlicher Ursachen	Kohortenstudien
Schätzung der Inzidenzrate in exponierten Bevölkerungen	ausschließlich Kohortenstudien
Untersuchung von Kofaktoren, die sich über die Zeit verändern	vorzugsweise Kohortenstudien
Schluss von Ursache auf Wirkung	Interventionsstudien

war die Responserate (gut ist $\geq 80\%$, ohne/geringe Aussagekraft $\leq 30\%$) und die Verlustrate, unter anderem durch Fortziehen und nachträgliche Teilnahmeverweigerung („loss to follow up“)? Um zu erkennen, ob sich Teilnehmer von Nichtteilnehmern unterscheiden, sollten auch Informationen zu den Nichtteilnehmern aufgeführt und diese mit denen der Teilnehmer verglichen werden. Durch Beschreibung der Auswahlkriterien und Verlustraten wird ersichtlich, ob die Stichprobe repräsentativ für die Zielpopulation ist. Zu einer guten Studienbeschreibung gehört auch die Erwähnung fehlender Werte (Missing-Werte). Insbesondere in Fall-Kontroll-Studien, aber auch in nicht randomisierten klinischen Studien und Kohortenstudien, sollte man die Wahl der Kontrollgruppe sehr genau beschreiben. So wird deutlich, dass die Kontrollgruppe mit der Studiengruppe vergleichbar ist und keine systematischen Abweichungen bestehen, die zu Verwechslungen („confounding“) beziehungsweise Störungen führen können (13).

Wird erklärt, wie Messungen durchgeführt wurden? Werden die eingesetzten Messinstrumente und die Messmethodik, wie zum Beispiel Messgeräte, Skala der gemessenen Werte, Labordaten und Zeitpunkt, ausreichend genau beschrieben? Erfolgt die Messungen unter standardisierten – und damit vergleichbaren – Bedingungen für alle Patienten? Die Angabe messspezifischer Informationen zur Genauigkeit (Reliabilität, Validität) ist wichtig. Der Leser muss erkennen, welches Skalenniveau die erhobenen Variablen haben (nominal zum Beispiel Augenfarbe, ordinal zum Beispiel Tumorstadium, metrisch zum Beispiel Körpergewicht), denn das Skalenniveau gibt vor, welche Auswertung möglich ist. Die deskriptive Analyse bedient sich beschreibender Maßzahlen und grafischer sowie tabellarischer Darstellungen, wohingegen bei der statistischen Auswertung die Wahl des Testverfahrens zu berücksichtigen ist. Die Interpretation und Aussagekraft der Ergebnisse wird ebenfalls vom Skalenniveau beeinflusst. Beispielsweise sollten ordinalskalierte Daten nicht

durch die Angabe von Mittelwerten beschrieben werden. Ging der Studiendurchführung eine sorgfältige Fallzahlplanung voraus? Wenn die Fallzahlen zu gering sind, kann möglicherweise ein tatsächlich vorhandener Unterschied, zum Beispiel in der Wirkung zweier Medikamente oder bei der Betrachtung eines Erkrankungsrisikos durch einen Umweltfaktor, nicht erkannt werden. Man spricht dann von zu niedriger Power.

Ergebnisse

Im Ergebnisteil sollen Untersuchungsergebnisse wertfrei, das heißt ohne Interpretation, und übersichtlich dargestellt werden. Die Interpretation der Resultate ist Aufgabe der anschließenden Diskussion. Die Ergebnisse sollten Antworten auf forschungsleitende Fragestellungen geben und gut strukturiert, verständlich sowie folgerichtig dargelegt werden. Zunächst sollten die Ergebnisse deskriptiv unter Angabe statistischer Kenngrößen, wie Fallzahlen, Mittelwerte, Streumaße und Konfidenzintervalle, dargestellt werden. Hierzu gehört auch die vollständige Beschreibung der Studienpopulation. Im zweiten, analytischen Teil werden Zusammenhänge zwischen Merkmalen beschrieben oder Risikomaße von bestimmten Einflussgrößen, zum Beispiel Rauchverhalten auf eine Zielgröße, wie Lungenkrebs, geschätzt und eventuell geeignete statistische Modelle gerechnet.

Neben der Angabe der statistischen Signifikanz in Form von p-Werten ist die umfassende deskriptive Beschreibung der Daten sowie die Angabe von Konfidenzintervallen und Effektstärken ausdrücklich zu empfehlen (14, 15, 16). Tabellen und Grafiken im Ergebnisteil dienen der Übersichtlichkeit, und die darin gegebenen Informationen sollen selbsterklärend sein.

Diskussion

Im Diskussionsteil sollte der Autor seine Ergebnisse ehrlich und offen diskutieren. Unabhängig vom Studientyp werden im Wesentlichen zwei Ziele verfolgt:

Vergleich der eigenen Ergebnisse mit dem aktuellen Stand der Wissenschaft – Folgende Fragen sollen im Diskussionsteil beantwortet werden: Welche neuen Erkenntnisse fügt die Studie dem bestehenden Kenntnisstand hinzu? Welche Schlussfolgerungen sind aus den Ergebnissen zu ziehen? Führt das Ergebnis der Studie zum Nach- beziehungsweise Überdenken der eigenen ärztlichen Tätigkeit, zum Beispiel zur Modifikation einer Behandlung oder der Beachtung bisher unbeachteter Aspekte? Geben die Ergebnisse Anlass zu weitergehenden Untersuchungen? Werden neue, unbeantwortete Fragen durch die Studie aufgeworfen? Was bedeuten die Resultate für die Wissenschaft, den klinischen Alltag, die Patientenversorgung und das ärztliche Handeln? Stehen die Resultate in Übereinstimmung mit dem Großteil der Ergebnisse anderer Studien? Falls nicht, welche Gründe könnte es dafür geben? Erscheinen die Ergebnisse aus biologischer beziehungsweise medizinischer Sicht plausibel?

Kritische Beschreibung der Grenzen der eigenen Ergebnisse – Könnten mögliche Fehlerquellen, ob zufälliger oder systematischer Art, die Ergebnisse beeinflusst haben? Selbst bei sehr sorgfältiger Studienplanung und

Durchführung können Fehler nie ganz ausgeschlossen werden. Vielleicht kam es im Verlauf der Untersuchung zu einem unerwartet hohen „loss to follow up“ (Verlustrate beispielsweise durch nachträgliche Teilnahmeverweigerung oder Fortzug). Bei Gruppenvergleichen stellt sich die Frage, ob sich aus der Beobachtung ausgeschiedene Teilnehmer zwischen den Gruppen unterscheiden. Möglicherweise konnte aus diesem Grund ein echter Unterschied zum Beispiel in einer Fall-Kontroll-Studie bezüglich eines Risikofaktors nicht erkannt werden. Das Auftreten eines Unterschieds aufgrund einer positiven Selektion der Studienpopulation ist gleichfalls möglich. Dieser Umstand und die Beschreibung der Patienten, die die Studie nicht abgeschlossen haben, muss in der Diskussion erwähnt werden. Eine mögliche Verzerrung der Studienergebnisse durch fehlende Werte sollte man ebenfalls diskutieren.

Systematische Fehler sind insbesondere in epidemiologischen Studien häufig, weil sie meist beobachtend und nicht experimentell kontrolliert sind. In Fall-Kontroll-Studien kann die rückblickende (retrospektive) Erfragung der Exposition von den Studienteilnehmern eine typische Fehlerquelle sein. Hier kann es zu erinnerungsbedingten Verzerrungen kommen („recall bias“). Bei Kohortenstudien ist eine häufige Fehlerquelle „confounding“ (Störgröße). Dieser Fehler tritt auf, wenn zwei Risikofaktoren in enger Beziehung zueinander stehen, aber beide zugleich auch mit der Zielgröße zusammenhängen. Solche Fehler können durch Adjustierung hinsichtlich der Störgröße korrigiert und aufgedeckt werden. Beispielsweise könnte in einer Untersuchung die Tatsache, dass Raucher überdurchschnittlich häufig Kaffee trinken, zu der irrtümlichen Annahme führen, dass Kaffeetrinken Lungenkrebs verursacht. Werden mögliche Confounder in der Publikation nicht berücksichtigt, sollte sich der kritische Leser fragen, ob die Ergebnisse nicht durch diese Fehlerquelle vorgetäuscht werden können. Falls man mögliche Confounder nicht in der Analyse berücksichtigt hat, sollten die Fehlermöglichkeiten zumindest kritisch erörtert werden. Weiterführende Literatur zu Fehlerquellen und Methoden der Korrektur bieten Beaglehole und auch Webb (17, 18).

Nicht signifikante Ergebnisse müssen ebenfalls publiziert werden. Leider werden signifikante Ergebnisse oft immer noch als höherwertig betrachtet, womit die Wahrscheinlichkeit für die Veröffentlichung signifikanter im Vergleich zu nicht signifikanten Ergebnissen steigt. Dieser Publikationsbias führt zu einer systematischen Verzerrung des wissenschaftlichen Kenntnisstandes. Nach einer aktuellen Übersichtsarbeit gilt das vor allem für klinische Studien (3). Nur bei Veröffentlichung aller validen Ergebnisse gut geplanter und korrekt durchgeführter Studien kann unverzerrt über den Einfluss eines Risikofaktors auf die Entstehung einer Krankheit, die Güte eines Diagnoseverfahrens, die Eigenschaften eines Werkstoffes oder den Erfolg einer Intervention, zum Beispiel einer Therapie, nachgedacht werden. Der Untersucher und das publizierende Journal haben somit die Verantwortung, ihren Teil dazu beizutragen, dass Entscheidungen über wichtige Fragestellungen unverfälscht unter Einbeziehung aller validen und wissenschaftlich fundierten Ergebnisse getroffen werden können.

KASTEN 2

Kritische Fragen

- Hat die Studie einen interessanten wissenschaftlichen Inhalt?
- Werden Aussagen und Zahlenwerte durch Literaturangaben belegt?
- Ist das Thema der Studie medizinisch relevant?
- Ist das Thema der Studie innovativ?
- Überprüft die Studie die formulierten Fragestellungen des Autors?
- Ist das Design der Studie geeignet, die Fragestellungen und/oder Hypothesen zu beantworten?
- Führten praktische Schwierigkeiten (beispielsweise bei der Rekrutierung, „loss to follow up“) zu schweren Kompromissen hinsichtlich der Durchführung gemäß dem Studienprotokoll?
- War die Anzahl fehlender Messwerte („missings“) für eine aussagekräftige Auswertung zu hoch?
- War die Fallzahl zu klein und somit die statistisch ermittelte Aussagekraft (Power) der Studie zu gering?
- War der Ablauf der Studie schlecht oder unzureichend kontrolliert (fehlende Messwerte, Confounding, Zeitverletzungen)?
- Ziehen die Autoren nicht gerechtfertigte Schlüsse aus den Ergebnissen?
- Gibt es unüberbrückbare Interessenkonflikte, bei denen die Autoren und/oder der Sponsor finanzielle und/oder ideologische Interessen haben?

Erwähnt sei hier der Unterschied zwischen statistischer Signifikanz, also der Minimierung des Zufalls, und klinischer Relevanz, was sich auf die Bedeutsamkeit des Ergebnisses bezieht. Durch die Wahl einer genügend großen Stichprobe können selbst sehr kleine Unterschiede statistisch signifikant werden. Damit ist aber nicht automatisch gesichert, dass es sich um relevante Ergebnisse handelt (13, 19). Das gilt sowohl für Ergebnisse epidemiologischer Studien aus Public-Health-Perspektive, als auch für klinische Studien aus klinischer Sicht. In beiden Fällen entscheidet letztendlich eine sorgfältige Kosten-Nutzen-Abwägung über Änderung oder Beibehaltung bestehender Praktiken. Auf Bevölkerungsebene stellt sich die Frage, wie häufig der untersuchte Risikofaktor tatsächlich auftritt und ob eine geringe Risikoerhöhung umfangreiche Public-Health-Interventionen rechtfertigt. Aus klinischer Sicht muss sorgfältig abgewogen werden, ob beispielsweise eine geringfügig bessere Wirkung eines neuen Präparates höhere Kosten und eventuell häufigere Nebenwirkungen rechtfertigt. Der Leser sollte den Unterschied zwischen statistischer Signifikanz und klinischer Relevanz erkennen, um die Ergebnisse richtig einschätzen zu können.

Schlussfolgerung

Bei den Schlussfolgerungen sollten sich die Autoren auf wenige, besonders wichtige Resultate beziehen. Eine zentrale Frage ist, ob die Interpretationen aus den Ergebnissen nachvollziehbar sind. Schlussfolgerungen, die weder durch die eigenen noch durch fremde Daten gestützt werden, sollte man vermeiden. Bei einer explorativen Datenanalyse von Beweisführung zu sprechen, ist nicht richtig.

TABELLE 2

Checkliste zur Beurteilung der Qualität wissenschaftlicher Veröffentlichungen

	ja	unklar	nein
Design			
Ist die eigentliche Fragestellung der Studie gut beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden die Studienpopulation(en) sowie Ein- und Ausschlusskriterien umfassend beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden die Patienten randomisiert auf die Studienarme verteilt? Falls ja: wurde die Methode der Randomisierung entsprechend beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
a) Wird zur Fallzahl Stellung genommen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Wurden ausreichend Fälle einbezogen (z. B. Power \geq 50 %)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
War die Messmethode (z. B. Labor, Fragebögen, Diagnosetest) geeignet hinsichtlich der Bestimmung der Zielgröße (bzgl. Skala, Untersuchungszeitpunkt, Standardisierung)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden Angaben zum Datenverlust (Responseraten, „loss to follow up“, „missing values“) gemacht?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Start und Durchführung der Studie			
Stimmen Behandlungs- und Kontrollgruppe hinsichtlich wichtiger studienrelevanter Merkmale (z. B. Alter, Geschlecht, Rauchgewohnheiten usw.) überein?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden die Drop-out-Fälle in Abhängigkeit von der Behandlungs- bzw. Kontrollgruppe deskriptiv ausgewertet?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wie viele Fälle standen während der gesamten Studiendauer unter Beobachtung?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden während der Studie auftretende Nebenwirkungen und unerwünschte Ereignisse beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analyse und Auswertung			
Waren die eingesetzten statistischen Größen und Methoden sinnvoll gewählt und ihre Beschreibung verständlich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden die statistischen Analyseverfahren verständlich beschrieben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden die wichtigen Einflussgrößen (prognostische Faktoren) berücksichtigt bzw. zumindest diskutiert?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
War die Präsentation der statistischen Größen angemessen, umfassend, übersichtlich und verständlich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wurden Effektstärken und Konfidenzintervalle zu den Hauptergebnissen angegeben?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wird ersichtlich, warum vorliegendes Studiendesign/statistische Methodik verwendet wurde?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sind alle Schlussfolgerungen durch die Ergebnisse der Studie gestützt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Anhand einer Checkliste ist es möglich, die statistische und methodische Qualität einer Studie einzuschätzen und bei Berücksichtigung der angesprochenen Punkte zu verbessern. Nicht alle aufgeführten Punkte eignen sich zur Beurteilung jedes Studientyps. Beispielsweise stellt sich die Frage nach einer Randomisierung vor allem bei klinischen Studien.

Auch bei konfirmatorischen Studien sollen die eigenen Ergebnisse zur Wahrung der Konsistenz immer im Kontext mit Resultaten anderer Studien gesehen werden. Bei der Beurteilung der Ergebnisse müssten auch Schwachstellen der Studie gebührend in die Schlussfolgerung einfließen. Nur unter Berücksichtigung der Möglichkeit fehlerhafter oder zufälliger Ergebnisse kann die Arbeit objektiv sein. Der Einbezug auch nicht signifikanter Ergebnisse trägt zur Glaubwürdigkeit bei. Hierbei soll „nicht signifikant“ nicht mit „es besteht kein Zusammenhang“ verwechselt werden. Signifikante Ergebnisse sollten unter dem Aspekt der biologischen und medizinischen Plausibilität betrachtet werden.

Sogenannte Evidenzskalen, wie sie etwa in einigen amerikanischen Zeitschriften verwendet werden, können den Leser hinsichtlich seiner Entscheidung, in welchem Maß das Gelesene in die Praxis einfließen soll, unterstützen (20). Solange nicht alle Zeitschriften solche Empfehlungen anbieten, ist weiterhin die Befähigung des einzelnen Arztes zum kritischen Lesen wissenschaftlicher Publikationen entscheidend, ob diagnostische und therapeutische Praxis auf aktuellem medizinischem Wissen basiert.

Literaturangaben

Die Zitierweise der Literaturangaben richtet sich nach den Vorgaben der Zeitschrift. In diesem Abschnitt müssen sich alle im Text, in den Tabellen und Grafiken zitierten Literaturstellen finden. Wichtig ist hierbei auf die Aktualität der Literaturstellen zu achten, damit klar wird, ob die Publikation neues Wissen einbezieht. Die zitierten Literaturstellen sollen für den Leser eine wertvolle Hilfe sein, um sich weiter in das Themengebiet zu vertiefen.

Danksagung und Interessenkonflikt

In diesem wichtigen Teil müssen sich Informationen zu Sponsoren der Studie finden. Es sollte klar ersichtlich sein, ob ein möglicher Interessenkonflikt, zum Beispiel finanzieller Art, besteht (21).

Zusammenfassend stellen *Tabelle 2* und *Kasten 2* die wesentlichen Punkte zur Beurteilung eines Artikels übersichtlich dar. Nicht alle Fragen dieser Checkliste treffen auf jede Veröffentlichung beziehungsweise jeden Studientyp zu. Wer sich weiter in die Thematik des Schreibens von wissenschaftlichen Veröffentlichungen

vertiefen möchte, sei auf Gardner et al. (19), Altman (7) sowie Altman et al. (22) verwiesen. Weitere Checklisten zur Beurteilung des statistischen Inhalts medizinischer Studien liefern Gardner et al. (23), Altman (7) und das CONSORT-Statement (12).

Interessenkonflikt

Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

Manuskriptdaten

eingereicht: 9. 1. 2007, revidierte Fassung angenommen: 7. 6. 2007

LITERATUR

- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, RW Scott: Evidence based medicine: what it is and what it isn't. Editorial. *BMJ* 1996; 312: 71–2.
- Albert DA: Deciding whether the conclusion of studies are justified: a review. *Med Decision Making* 1981; 1: 265–75.
- Gluud LL: Bias in clinical intervention research. *Am J Epidemiol* 2006; 163: 493–501.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. *Lancet* 1991; 337: 867–72.
- Greenhalgh T: How to read a paper: getting your bearings (deciding what the paper is about). *BMJ* 1997; 315: 243–6.
- Kallet H: How to write the methods section of a research paper. *Respir Care* 2004; 49: 1229–32.
- Altman DG: Practical statistics for medical research. London: Chapman and Hall 1991.
- De Simonian R, Charette LJ, Mc Peek B, Mosteller F: Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1332–7.
- Trampisch HJ, Windeler J, Ehle B: Medizinische Statistik. Berlin, Heidelberg, New York: Springer 2000, 2. überarb. Auflage.
- Klug SJ, Bender R, Blettner M, Lange S: Wichtige epidemiologische Studientypen. *Dtsch Med Wochenschr* 2004; 129: T7–T10.
- Schumacher M, Schulgen G: Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung. Berlin: Springer 2006, 2. überarb. Auflage.
- Moher D, Schulz KF, Altman DG für die CONSORT Gruppe: Das CONSORT Statement: Überarbeitete Empfehlungen zur Qualitätsverbesserung von Reports randomisierter Studien im Parallel-Design. *Dtsch Med Wochenschr* 2004; 129: T16–T20.
- Blettner M, Heuer C, Razum O: Critical reading of epidemiological papers. A guide. *J Public Health* 2001; 11: 97–101.
- Borenstein M: The case for confidence intervals in controlled clinical trials. *Control Clin Trials* 1994; 15: 411–28.
- Gardner MJ, Altman DG: Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* 1986; 292: 746–50.
- Bortz J, Liener GA: Kurzgefaßte Statistik für die Klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben. Berlin: Springer 2003; 2. Auflage, 39–45.
- Beaglehole R, Bonita R, Kjellström T: Einführung in die Epidemiologie. Bern, Göttingen, Toronto, Seattle: Huber 1997.

- Webb P, Bain C, Pirozzo S: Essential epidemiology. An introduction for students and health professionals. New York: Cambridge University Press 2005.
- Gardner MJ, Machin D, Brynart TN, Altman DG: Statistics with confidence. Confidence intervals and statistical guidelines. London: BMJ books 2002.
- Ebell MH, Siwek J, Weiss BD et al.: Simplifying the language of evidence to improve patient care. *J Fam Pract* 2004; 53: 111–20.
- Bero LA, Rennie D: Influences on the quality of published drug studies. *Int J Technol Assess Health Care* 1996; 12: 209–37.
- Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical guidelines for contributors to medical journals. *BMJ* 1983; 286: 1489–93.
- Gardner MJ, Machin D, Campbell MJ: Use of check lists in assessing the statistical content of medical studies. *BMJ (Clin Res Ed)* 1986; 292: 810–2.

Anschrift für die Verfasser

Dr. med. Jean-Baptist du Prel, MPH
 Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
 Universitätsklinikum Mainz
 Obere Zahlbacher Straße 69, 55131 Mainz
 E-Mail: duprel@imbei.uni-mainz.de

SUMMARY

Critical Appraisal of Scientific Articles—Part 1 of a Series on Evaluation of Scientific Publications

Introduction: In the era of evidence-based medicine, one of the most important skills a physician needs is the ability to analyze scientific literature critically. This is necessary to keep medical knowledge up to date and to ensure optimal patient care. The aim of this paper is to present an accessible introduction into critical appraisal of scientific articles.

Methods: Using a selection of international literature, the reader is introduced to the principles of critical reading of scientific articles in medicine. For the sake of conciseness, detailed description of statistical methods is omitted.

Results: Widely accepted principles for critically appraising scientific articles are outlined. Basic knowledge of study design, structuring of an article, the role of different sections, of statistical presentations as well as sources of error and limitation are presented. The reader does not require extensive methodological knowledge. As far as necessary for critical appraisal of scientific articles, differences in research areas like epidemiology, clinical, and basic research are outlined. Further useful references are presented.

Conclusion: Basic methodological knowledge is required to select and interpret scientific articles correctly.

Dtsch Arztebl Int 2009; 106(7): 100–5
 DOI: 10.3238/arztebl.2009.0100

Key words: publication, critical appraisal, decision making, quality assurance, study



The English version of this article is available online:
www.aerzteblatt-international.de