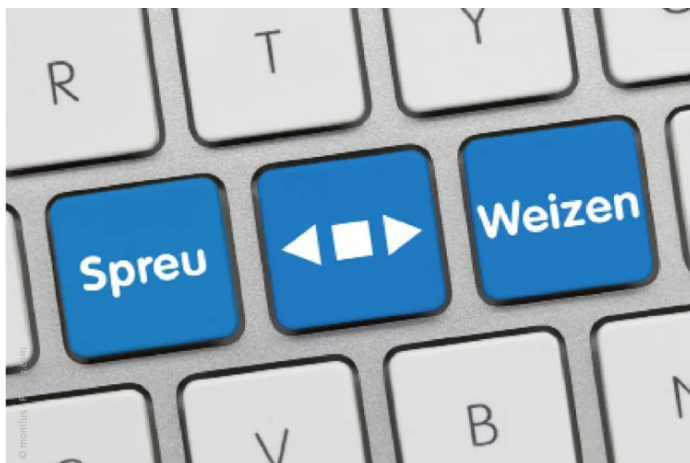


01.03.2014 Qualitätsmessung

Praxis II: Critical Appraisal unter Zuhilfenahme von Instrumenten der Qualitätsbeurteilung – oder „Die Spreu vom Weizen trennen“


F. Pianka, U. Klaiber, P. Knebel, M. K. Diener



Die Evidenzbasierte Medizin (EbM) basiert auf einer Zusammensetzung dreier essenziellen Zutaten: dem Patienten und dessen Bedürfnisse, dem Arzt und dessen klinische Expertise und der Gesamtheit der verfügbaren relevanten Evidenz. Die Schnittmenge dieser, also eine patientenorientierte Behandlung die auf klinischer Expertise und empirischer Evidenz beruht, nennt man Evidenzbasierte Medizin (**Abb. 1**). Der folgende Artikel befasst sich mit der externen Evidenz und deren


Beurteilung durch Zuhilfenahme von Instrumenten der Qualitätsbeurteilung und beleuchtet außerdem ihren Stellenwert in der Bewertung chirurgischer Studien.

Abb. 1: Zusammensetzung der Evidenzbasierten Medizin (EbM)

 OEBPS/images/02_04_A_03_2014_PraxisII_image_01.jpg

Um in der Masse der vorhandenen Evidenz die entsprechenden Studien für die persönliche Evidenzsynthese oder Entscheidungsfindung zu identifizieren, ist eine Beurteilung der relevanten Studien nach kritischen Gesichtspunkten im Sinne eines *Critical Appraisal* unumgänglich. 2013 waren in der online Literatur- Suchmaschine *Pubmed.com* über eine Millionen Publikationen verzeichnet. Der Anteil klinischer Studien als mögliche relevante Grundlage für die klinische Entscheidungsfindung betrug jedoch nur knapp 10 % (**Abb. 2**).

Abb. 2: Anteil klinischer Forschung an der Gesamtheit wissenschaftlicher Publikationen 2013 (Datenquelle: www.pubmed.com)

 OEBPS/images/02_04_A_03_2014_PraxisII_image_02.jpg

Unter diesem Gesichtspunkt die vergleichbar geringe Anzahl an klinisch relevanter Literatur zu kategorisieren und Studien hoher Qualität von niedriger Qualität zu unterscheiden, werden zunehmend Instrumente der Qualitätsbeurteilung entwickelt, um eine systematische und standardisierte Bewertung zu ermöglichen. Letztere ist essentiell um die Wertigkeit der jeweiligen Evidenz für deren klinische Anwendung zu überprüfen. Die Evidenz aus kontrollierten klinischen Studien, systematischen Übersichtsarbeiten und Meta-Analysen nimmt dabei eine zentrale Rolle ein.

Die Qualitätsbeurteilung basiert in der Regel auf den so genannten *Levels of Evidence*, der hierarchischen Anordnung von Studientypen und -designs bezogen auf den Wert der wissenschaftlichen Aussagekraft. Diese wird, entsprechend methodischer Charakteristika, anhand der studienbezogenen *internen Validität* beurteilt (siehe www.CEBM.net, Oxford 2001 „*Levels of Evidence*“).

Eine gut begründete Studienbewertung ist für die Anwendung im klinischen Alltag ebenso bedeutsam wie für die Entscheidungen auf gesundheitspolitischer Ebene. Um Bedeutung und Relevanz von Studienergebnissen adäquat zu interpretieren sowie für die Praxis anwendbar zu machen, bieten sich sowohl für den Leser als auch den Anwender von klinischer Evidenz diverse Instrumente der Qualitätsbeurteilung an.

Die Qualitätsbeurteilung

Begriff der „Qualität“ und Stellenwert für die Interpretation klinischer Studien

Die Qualität einer klinischen Studie bezieht sich im Wesentlichen auf die Aspekte Design, Durchführung und Analyse, die sogenannte *interne Validität*. Diese beschreibt die Wahrscheinlichkeit, dass die Ergebnisse der Studie den „echten“ Effekt einer therapeutischen Intervention widerspiegeln. Diese Wahrscheinlichkeit lässt sich durch die Beurteilung von Zufallsfehlern, Bias und Störfaktoren abschätzen und nimmt in der Regel mit steigendem Evidenzlevel zu. Zufällige und/oder systematische Fehler (z. B. selection bias, measurement bias) sowie Confounding verzerren das Studienergebnis und reduzieren die *interne Validität*. Die Qualitätsbeurteilung kann diese und das Potenzial möglicher Verzerrungen messen und erlaubt somit Rückschlüsse auf die Qualität der Planung, Durchführung und Analyse einer Studie.

Der Begriff der *externen Validität* beschreibt die Generalisierbarkeit sowie die Anwendbarkeit der Studienergebnisse auf die Zielpopulation. Qualität bedeutet somit im Idealfall eine methodologisch einwandfreie Studienidee mit klinisch-relevanten sowie patientenorientierten Endpunkten, die nach valider Prüfung eine Entscheidungsfindung erlauben und eine tatsächliche Verbesserung der Patientenversorgung ermöglichen.

Kernelemente der Qualitätsbeurteilung

Die kritische Beurteilung einer wissenschaftlichen Publikation sollte mindestens folgende drei Leitfragen beinhalten:

a. Sind die Ergebnisse der Studie valide?

Die Beantwortung dieser Frage bezieht sich auf die *interne Validität* der zu beurteilenden Studie. Hier werden vor allem der Randomisationsprozess, die Verblindung der Patienten und/oder des Studienpersonals und die Standardisierung der chirurgischen Interventionen und des perioperativen Managements beurteilt, um die methodische Robustheit der Studie zu erfassen.

b. Wie sind die Ergebnisse und Therapieeffekte?

Wenn die Ergebnisse in einem ersten Schritt als valide und methodisch korrekt erfasst beurteilt wurden, sollte im zweiten Schritt die Größe und Präzision des beobachteten Therapieeffektes bestimmt werden. Da der „wahre Unterschied“ nie eindeutig bestimmt werden kann, sollten die Punktschätzer und Präzisionsmaße detailliert evaluiert werden, bevor eine klinische Empfehlung daraus entwickelt wird. Am besten geeignet für eine Einschätzung der Präzision sind 95 % Konfidenzintervalle, welche wesentlich aussagekräftiger sind als alleinige P-Werte (Signifikanzwert) [1].

c. Helfen die Ergebnisse bei der Entscheidungsfindung?

Nach Erfassung der *internen Validität* und der Präzision eines Studienergebnisses gilt es, den klinischen Stellenwert zu beurteilen. In diesem Zusammenhang werden oftmals die Begriffe *externe Validität* und *Generalisierbarkeit* gebraucht. Die Relevanz des Studienergebnisses beruht maßgeblich auf dieser externen Validität, was wiederum die Voraussetzung für die Anwendbarkeit auf eine definierte Patientengruppe in einer bestimmten klinischen Situation ist [2].

Die Methodik des *Critical Appraisal* kann gezielt erlernt werden und hilft dem Anwender irreführende Ergebnisse und ungenaue Datenpräsentation aufzudecken. Im Folgenden werden die gängigsten Instrumente zur Qualitätsbeurteilung von klinischen Studien aufgeführt und erläutert.

Häufig eingesetzte Instrumente zur Qualitätsbeurteilung

a. Das CONSORT – Statement

Da in einer Vielzahl an Studienpublikationen sehr häufig wichtige Elemente der Methodik und Auswertung lückenhaft berichtet wurden [3], wurde Mitte der 1990er Jahre das CONSORT-Statement (Consolidated Standards of Reporting Trials) entwickelt.

Nach Untersuchungen von Solomon et al. [4] wiesen besonders chirurgische Studien meist einen sehr knapp gehaltenen Methodenteil auf, es fehlte häufig eine Fallzahlplanung sowie adäquate Definitionen von Endpunkten, was für die Vergleichbarkeit von Studienergebnissen unerlässlich ist.

Zur Verbesserung der Berichtsqualität können Checklisten nach dem Vorbild des CONSORT-Statements über das Erfassen wesentlicher methodischer sowie inhaltlicher Kriterien hilfreich sein. **Tabelle 1** gibt einen Überblick der wichtigsten Checklisten aus dem Bereich der klinischen Forschung, die unter den angegebenen Referenzen kostenfrei abrufbar sind.

b. GRADE

Der GRADE-Ansatz unterscheidet zwischen der Qualitätsbeurteilung im Rahmen systemischer Übersichtsarbeiten und der Qualitätsbeurteilung als Baustein der Leitlinienentwicklung. Weiterhin unterscheidet GRADE zwischen Qualität der Evidenz und Empfehlungsstärke. Kernelement ist das PICO-Schema (Patientenkollektiv, Intervention, Vergleich/Kontrolle, Outcomes), welches als Ausgangspunkt für die Qualitätsbewertung der Evidenz dient. Im Vorfeld wird festgelegt, welche Endpunkte berücksichtigt bzw. nicht berücksichtigt werden. Diese werden dabei nach Relevanz für eine Empfehlung gewichtet (essenziell, wichtig aber nicht essenziell, von begrenzter Bedeutung). Die

Qualität kann dabei für verschiedene Endpunkte unterschiedlich sein. Die verschiedenen Level der Qualität bei GRADE sind dabei wie folgt unterteilt: hoch, moderat, niedrig, sehr niedrig [5].

c. Cochrane Risk of Bias Tool (CCRBT)

Seit seiner Einführung im Jahr 2008 wird das Cochrane Risk of Bias Tool von der Cochrane Collaboration stetig weiterentwickelt, um die Nachteile bestehender Instrumente der Qualitätsbeurteilung zu identifizieren und diesen zu begegnen. Von Bedeutung ist zudem, die Unterschiede zwischen tatsächlicher Studiendurchführung und der Berichterstattung/ wissenschaftlichen Publikation aufzudecken.

Der Ansatz des CCRBT begründet sich auf der individuellen Einschätzung des Verzerrungspotenzials (hoch, niedrig, unklar) von sechs Schlüsselkomponenten einer Studie (Generierung der zufälligen Zuteilung, verdeckte Zuteilung, Verblindung, unvollständiges Berichten der Endpunkte, selektives Berichten der Endpunkte, andere Möglichkeiten von Verzerrung). Jede Komponente wird separat bewertet und lässt somit eine zuverlässige Einschätzung über den Einfluss dieser Komponente auf den Behandlungseffekt zu [6]. Die Bewertung erfolgt dabei im Idealfall nicht allein auf Basis der Publikation, sondern ebenfalls auf weiteren studienbezogenen Dokumenten, wie z. B. auf dem Studienprotokoll.

d. Die CLEAR NPT Checkliste

Die Clear NPT (Non Pharmacological Treatments) Checkliste ist ein von 55 Experten (Biometriker, Epidemiologen, Kliniker) entwickeltes Instrument zur Beurteilung von nicht-pharmakologischen randomisiert kontrollierten Studien [7]. Insgesamt umfasst die Checkliste zehn Elemente mit je fünf Unterelementen. Unter anderem beziehen sich die ausgewählten Elemente auf die Standardisierung der Intervention, den möglichen Einfluss des medizinischen Personals sowie auf zusätzliche Maßnahmen zur Minimierung einer potenziellen Verzerrung durch fehlende Verblindung. Ein detaillierter Leitfaden als Hilfestellung zur Beurteilung der jeweiligen Elemente ist unter <http://www.bichat.inserm.fr/equipes/Emi0357/docs/usersguidelines.pdf> öffentlich verfügbar. Die Vorteile der CLEAR NPT Checkliste für die Beurteilung randomisiert kontrollierter Studien in der Chirurgie ergeben sich aus der Charakteristik und Gewichtung chirurgisch relevanter Kriterien, welche durch andere Instrumente nicht abgedeckt werden.

Tab. 1: Einschätzung des Nutzen Evidenzbasierter Medizin

Akronym	Ausführlicher Name	Studientyp	Referenz
CONSORT	Consolidated Standards of Reporting Trials	RCT	Altman et al. [10] www.consort-statement.org
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology	Kohorten-/ Fall-Kontroll-/ Querschnittstudien	Vandenbroucke et al. [11]
MOOSE	Meta-Analysis of Observational Studies in Epidemiology	Meta-Analyse von Beobachtungsstudien	Stroup et al. [12]

QUOROM	Quality of Reporting of Meta-Analyses	Meta-Analyse von RCT	Moher et al. [13]
PRISMA (Weiterentwicklung von QUOROM)	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	Meta-Analyse von Studien zu Interventionen im Gesundheitswesen	Liberati et al. [14] www.prisma-statement.org
QUADAS	Quality Assessment of Studies of Diagnostic Accuracy included in systematic reviews	Diagnostik Studien	Whiting et al. [15]
STARD	Strengthening the Reporting of diagnostic	Diagnostik Studien	Bossuyt PM et al. [16] www.stard-statement.org

Besonderheiten der Qualitätsbeurteilung chirurgischer Studien

Die Chirurgie hat sich im Rahmen der Verbreitung von EbM über die letzten 20 Jahre analog der pharmakologischen Forschung deutlich weiterentwickeln müssen. Die Untersuchung von Wirksamkeit und Sicherheit chirurgischer Verfahren musste nun ebenfalls anhand klar definierter Gesichtspunkte durchgeführt und evaluiert werden. Die in diesem Zuge entstandenen chirurgisch gewichteten Instrumente der Qualitätsbeurteilung unterschieden sich klar von denen pharmakologischer Studien (*Weiterführende Literatur: die IDEAL-Empfehlungen [8]*).

Chirurgische Interventionen sind meist komplex, eine Beurteilung erfordert daher eine detaillierte Beschreibung der chirurgischen Intervention [9]. Dazu gehören neben den technischen Details des Eingriffes auch das prä-, intra- und postoperative Management inklusive Anästhesie und Follow-Up. Die Expertise der durchführenden Chirurgen ist ebenfalls ein potenzieller Einflussfaktor für das Studienergebnis, obgleich sich dieser häufig als schwer evaluierbar herausstellt.

Im Vergleich zu Medikamentenstudien mit meist standardisierter Intervention, sind bei der Qualitätsbeurteilung chirurgischer Studien zahlreiche zusätzliche Kriterien von Bedeutung. Neben Lernkurve und Erfahrung der ausführenden Chirurgen gehören dazu auch Timing der Studiendurchführung und Standardisierung der chirurgischen Interventionen in allen Behandlungs- und Kontrollgruppen. Diese weiteren Parameter haben neben der reinen Studienmethodik einen erheblichen Einfluss auf das Studienergebnis und damit auch auf die Qualität und sollten deshalb ebenfalls erfasst und analysiert werden.

Schlussfolgernd sind allgemeine Instrumente der Qualitätsbeurteilung für chirurgische Studien häufig nicht ausreichend und sollten je nach Fragestellung und Fachgebiet erweitert und angepasst werden.

Fazit

Um der Flut an wissenschaftlichen Publikationen und stets steigender Verfügbarkeit (z. B. MEDLINE, EMBASE, Cochrane Library) kritisch begegnen zu können, bedarf es einer gewissen Expertise. Um das Herausheben der entscheidenden Studien zu erleichtern sind mittlerweile verschiedene Hilfsmittel und Instrumente der Qualitätsbeurteilung unerlässlich und erlauben ein Vorgehen nach standardisierten und bereits validierten Verfahren.

Nur qualitativ hochwertige Studien ermöglichen eine fundierte Entscheidungsfindung unter Einbezug aller notwendigen Kriterien. Patientenorientierung, Wirtschaftlichkeit und medizinische Anwendbarkeit können je nach Gewichtung der persönlichen Präferenz eine größere oder kleinere Rolle in der Suche nach Evidenz spielen. Um diese im Rahmen eines *Critical Appraisal* zu bewerten kann der medizinische Anwender bereits aus einer Vielzahl an Instrumenten der Qualitätsbeurteilung wählen, bisher jedoch ist eine Empfehlung im Sinne eines „one fits all“-Tools nicht möglich. Durch eine individuelle Anpassung an Fachgebiet, Fragestellung und Studientyp ist allerdings im Gegensatz zu allgemeinen Methoden meist sogar eine qualitativ hochwertigere Literaturbewertung durchzuführen. Somit sind Instrumente der Qualitätsbeurteilung im Zeitalter des Informationsüberflusses ein nicht mehr wegzudenkendes Mittel, die „Spreu vom Weizen“ zu trennen und hochqualitative Evidenz zu synthetisieren und daraus hervorgehende Evidenzbasierte Entscheidungen in der modernen Patientenversorgung zu treffen.

Literatur

- [1] Altman, D.G., *Confidence intervals for the number needed to treat*. BMJ, 1998. 317(7168): p. 1309-12.
- [2] Rothwell, P.M., *External validity of randomised controlled trials: "to whom do the results of this trial apply?"*. Lancet, 2005. 365(9453): p. 82-93.
- [3] Schulz, K.F., et al., *Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals*. JAMA, 1994. 272(2): p. 125-8.
- [4] Solomon, M.J., et al., *Randomized controlled trials in surgery*. Surgery, 1994. 115(6): p. 707-12.
- [5] Guyatt, G.H., et al., *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations*. BMJ, 2008. 336(7650): p. 924-6.
- [6] Higgins, J.P. and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. 2011.
- [7] Boutron, I., et al., *A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus*. J Clin Epidemiol, 2005. 58(12): p. 1233-40.
- [8] McCulloch, P., et al., *No surgical innovation without evaluation: the IDEAL recommendations*. Lancet, 2009. 374(9695): p. 1105-12.
- [9] Ergina, P.L., et al., *Challenges in evaluating surgical innovation*. Lancet, 2009. 374(9695): p. 1097-104.
- [10] Altman, D.G., *Better reporting of randomised controlled trials: the CONSORT statement*. BMJ, 1996. 313(7057): p. 570-1.

- [11] Vandembroucke, J.P., et al., *Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration*. PLoS Med, 2007. 4(10): p. e297.
- [12] Stroup, D.F., et al., *Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group*. JAMA, 2000. 283(15): p. 2008-12.
- [13] Moher, D., et al., *Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses*. Lancet, 1999. 354(9193): p. 1896-900.
- [14] Liberati, A., et al., *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration*. PLoS Med, 2009. 6(7): p. e1000100.
- [15] Whiting, P., et al., *The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews*. BMC Med Res Methodol, 2003. 3: p. 25.
- [16] Bossuyt, P.M., et al., *Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative*. Am J Clin Pathol, 2003. 119(1): p. 18-22.

Weiterführende Informationen	
Tipps zur Literaturbewertung	
Checklisten, sowie Critical Appraisal Software zum Download	
Checklisten zur Literaturbewertung	
Critical Appraisal Skills Programme, Literatur und Checklisten	
AGREE – englischsprachiges Tool zur Leitlinienbewertung	
DELBI – Deutsches Instrument zur Leitlinienbewertung	

Pianka F. / Klaiber U. / Knebel P. / Diener M. K. *Praxis II: Critical Appraisal unter Zuhilfenahme von Instrumenten der Qualitätsbeurteilung – oder „Die Spreu vom Weizen trennen“*. Passion Chirurgie. 2014 März, 4(03): Artikel 02_04.