

P. Kranke, R. Müllenbach, F. Schuster,  
A. Redel, M. Lange, E.-M. Kranke<sup>1</sup>,  
N. Roewer, L. H. J. Eberhart<sup>2</sup>

Klinik für Anaesthesiologie  
der Universität Würzburg  
(Direktor: Prof. Dr. Norbert Roewer)

<sup>1</sup>Klinik für Thorax-, Herz- und  
Thorakale Gefäßchirurgie  
der Universität Würzburg  
(Direktor: Prof. Dr. Rainer G. Leyh)

<sup>2</sup>Klinik für Anaesthesiologie der  
Universität Marburg  
(Direktor: Prof. Dr. Hinnerk Wulf)

# Systematische Übersichtsarbeiten und Metaanalysen in der klinischen Medizin – Erstellung, Bedeutung und Interpretation sowie Abgrenzung zur klassischen Übersichtsarbeit

## ZUSAMMENFASSUNG

Originalarbeiten dienen in erster Linie der Darstellung neuer wissenschaftlicher Ergebnisse und sind daher nur bedingt geeignet, bereits vorhandene Erkenntnisse auf einem bestimmten Gebiet für den Leser übersichtlich und didaktisch gut aufbereitet zu präsentieren. Diese Lücke füllen Übersichtsarbeiten oder Weiterbildungsartikel, die von vielen Klinikern zur Vertiefung und Aktualisierung von Lehrbuchinhalten geschätzt werden. Seit nunmehr geraumer Zeit wird die Lektüre solcher Beiträge und die Bearbeitung darauf abgestimmter Fragenkataloge z. B. von Ärztekammern als Qualifikationsnachweis für die ärztliche Weiterbildung anerkannt.

Allerdings weisen auch diese klassischen Übersichtsartikel einige Limitationen auf, die im vorliegenden Beitrag näher dargestellt werden. Diese betreffen in erster Linie die subjektive und für den Leser nicht nachvollziehbare Auswahl und Interpretation der Primärliteratur. Systematische Übersichtsarbeiten unterscheiden sich in diesem Punkt wesentlich von den klassischen Übersichtsartikeln, da sie in der Regel wie eine Originalarbeit aufgebaut sind. Sie beinhalten eine exakte Fragestellung, ein Studienprotokoll, eine umfassende Datengewinnung und deren statistische Synthese (Metaanalyse). Umfasst eine systematische Übersicht ebenfalls eine Metaanalyse, spricht man auch von einer quantitativen systematischen Übersicht. Häufig gelingt es damit, eine nur schwer überschaubare Flut von Veröffentlichungen zu einem Thema und auch Informationen aus schwer zugänglichen oder fremdsprachlichen Journalen auf eine für den Leser klar verständliche Aussage zu reduzieren.

Ziel der vorliegenden Arbeit ist es, die einzelnen Schritte einer solchen systematischen Übersicht zu erläutern und einige für die

se Publikationsart typische Begriffe, Abbildungstypen und Maßzahlen anhand von Beispielen zu erklären. Darüber hinaus werden Suchstrategien vorgestellt, die das Auffinden von systematischen Übersichtsarbeiten erleichtern. Aber auch die Schwachstellen und Limitationen dieser Publikationsform sollen aufgezeigt und dem Leser Hilfen für die kritische Bewertung solcher Arbeiten an die Hand gegeben werden.

## SCHLÜSSELWÖRTER

Statistische Verfahren: systematische Übersicht, quantitative systematische Übersicht, Metaanalyse, Review

## ABSTRACT

While original papers deal with new research results, they are unlikely to provide readers with a didactic well-structured background information on a specific topic. This gap is filled by review articles which are popular tools to inform clinicians in addition to textbook articles. The value of review articles has been accredited by the fact that they became part of the continuous medical education.

However, the limitations of traditional reviews are evident. They are valuable to obtain a comprehensive overview including all aspects of a topic while relevant efficacy data with respect to specific interventions usually are lacking. On the other hand systematic reviews summarise all available scientific data on a narrow and focused clinical question and – similar to controlled trials – follow well-defined guidelines. Metaanalyses summarise the findings of a systematic review using statistical methods. As they usually allow to obtain a precise and comprehensive quantification of the efficacy for specific interventions, they became the most important tool of evidence-based medicine. Thus, an initial overwhelming

quantity of literature including foreign language publications can be condensed to a clinically relevant statement that might be transferred into daily practice.

The aim of this paper is to guide through a sample metaanalysis. The focus is on the methodology of systematic reviews starting with the initial clinical question up to the graphical visualisation of the results. In addition, limitations of systematic reviews, search strategies to find systematic reviews and a guide to assess their validity will be discussed.

## KEY WORDS

Statistics: systematic review, quantitative systematic review, metaanalysis, review

## EINLEITUNG

**Die Fülle an medizinischen Fachpublikationen ist für den Einzelnen nicht mehr überschaubar.**

Die Flut an neuen Publikationen ist in vielen Bereichen der Medizin kaum mehr zu übersehen. Selbst wenn man sich nur auf die Arbeiten aus der klinischen Forschung beschränken würde, könnte man mit der immer weiter zunehmenden Publikationszahl nicht mehr Schritt halten. Für den Bereich der Inneren Medizin z. B. bedürfte es unter Berücksichtigung der zehn führenden Zeitschriften eines Leseaufwands von 19 Artikeln pro Tag, um auf dem aktuellen Stand der Wissenschaft zu bleiben [1]. In den Bereichen Kardiologie, Kardiotechnik oder Kardiotechnik dürften die Zahlen nicht wesentlich anders aussehen, wenn weltweit bereits im Jahre 1992 mehr als 50 Anästhesiezeitschriften lokalisiert werden konnten [2].

**Originalarbeiten, die „zufällig“ (unsystematisch) gelesen werden, können zu Verwirrung und Frustration beim Leser führen.** In der Praxis erfolgt das Überfliegen abonniertes oder in der Klinikbibliothek auslie-

gender Zeitschriften in der Regel wenig selektiert. Oft bleibt es dem Zufall überlassen, bei welcher Arbeit der Leser „hängen bleibt“. Bei Originalarbeiten kommt es vorrangig auf die exakte Darstellung der Ergebnisse und der verwendeten Methodik an. Schon allein aus Platzgründen können Einleitung und Diskussionsabschnitt das behandelte Thema und die klinischen Hintergründe nur ungenügend abhandeln. Raum für eine wirklich umfassende Darstellung der bislang zum behandelten Thema publizierten Arbeiten bleibt nur in den seltensten Fällen.

Erschwerend kommt hinzu, dass sich die Ergebnisse verschiedener Studien zu ein und demselben Thema nicht selten konträr gegenüberstehen. So wertvoll eine kontrovers geführte Diskussion in der Literatur für die Fachleute auf einem bestimmten Gebiet sein mag, so führt sie beim Leser, der in seiner knapp bemessenen Fortbildungszeit schnell verfügbare Informationen und Handlungsstrategien für seine tägliche Praxis sucht, doch eher zu Verwirrung, Frustration und Verärgerung.

### ***Traditionelle Übersichtsarbeiten bieten einen schnellen und umfassenden Überblick über ein Themengebiet.***

In diesem Dilemma ist es daher naheliegend, auf entsprechende Übersichtsartikel zurückzugreifen. Die sogenannten „narrative reviews“ (narrative = erzählend) sind die klassische Form der Übersichtsarbeit. Unter diese Rubrik fallen auch Fortbildungsartikel (educational reviews) in den entsprechenden Fachzeitschriften oder „refresher courses“, die Inhalte von Hauptvorträgen auf wissenschaftlichen Kongressen komprimiert wiedergeben. Da häufig auch Grundlagen aus Anatomie, Pathophysiologie oder Pharmakologie wiederholt werden, stellen sie eine ideale Ergänzung und Aktualisierung zu Lehrbüchern dar. Nicht ohne Grund werden die entsprechenden Weiterbildungsartikel und die Bearbeitung darauf abgestimmter Fragenkataloge auch von den Ärztekammern als Qualifikationsnachweis für die ärztliche Weiterbildung anerkannt [3].

### ***Traditionelle Übersichtsarbeiten weisen Limitationen auf.***

Die Autoren dieser Arbeiten sind in aller Regel Wissenschaftler, die sich schon längere Zeit mit dem entsprechenden Thema auseinandergesetzt haben. Somit ist es nahezu unvermeidlich, dass beim Abfassen des Manuskripts – bewusst oder unbewusst – auch eigene Meinungen mit einfließen. Weiterhin besteht die Tendenz, eigene Er-

gebnisse oder die Publikationen befreundeter Arbeitsgruppen überproportional häufig zu zitieren, gemäß dem Motto „Zitierst du mich, zitier ich dich“ [4], was die Gefahr in sich birgt, Sachverhalte über- oder unterzubewerten. Somit ist ein gewisses Maß an Subjektivität unvermeidlich. Auch wenn die Aussagen und Argumente im Text jeweils sorgfältig mit Literaturquellen belegt sind, wurden möglicherweise andere Arbeiten mit konträren Ergebnissen überhaupt nicht erwähnt oder eine Aussage wird durch eine „Scheingewissheit“ untermauert, indem wiederum populäre Übersichtsarbeiten zitiert werden, wie dies zum Beispiel für die scheinbar plausible positive Korrelation von erhöhtem Body-Mass-Index und der Inzidenz von Übelkeit und Erbrechen in der postoperativen Phase aufgezeigt werden konnte [5]. Die Auswahl der zitierten Primärliteratur ist in aller Regel nicht nachvollziehbar und der Leser muss auf die Objektivität des Verfassers der Übersichtsarbeit vertrauen, dem als „Meinungsbildner“ auf dem bearbeiteten Gebiet die nötige Kompetenz zugeschrieben wird.

## **SYSTEMATISCHE ÜBERSICHTEN**

### ***Unterschiede zwischen traditionellen Übersichtsarbeiten und systematischen Übersichten.***

Der Autor einer klassischen Übersichtsarbeit unterwirft sich lediglich den Formatierungsregeln (Umfangsbeschränkungen von Text und Zitaten) des jeweiligen Journals. Wie er dagegen die inhaltliche Gestaltung des Textes vornimmt und welche Schwerpunkte er dabei setzt, wird ihm in der Regel nicht vorgeschrieben. Die Lesbarkeit wird durch den erzählenden Charakter gefördert, da kein methodisches Hintergrundwissen zum Verständnis nötig ist.

Im Gegensatz dazu wird eine systematische Übersichtsarbeit nach einem vordefinierten Protokoll durchgeführt. Dazu hat mittlerweile bereits eine Expertenkommission konkrete Empfehlungen herausgegeben [6], die teilweise – ähnlich wie das sogenannte CONSORT-Statement für kontrollierte klinische Studien [7] – bereits zur Bedingung für die Annahme von systematischen Übersichten in Anästhesiejournalen avanciert ist [8]. Dies soll einerseits gewisse Standards gewährleisten, aber andererseits auch im Zuge einer Vereinheitlichung für bessere Lesbarkeit der klinischen Botschaft in den systematischen Übersichten sorgen. Anders als bei der klassischen Übersichtsarbeit, die sich weitgehend von selbst erschließt, finden sich in der Metho-

dik einer systematischen Übersicht häufig wiederkehrende – auf den ersten Blick möglicherweise abschreckende – methodische Vorgehensweisen, die nicht zwangsweise selbsterklärend sind. Im Folgenden soll deshalb der Entstehungsprozess einer systematischen Übersicht Schritt für Schritt erläutert werden.

### **Fragestellung**

Während eine normale Übersicht häufig ein großes Thema abhandelt und dabei auch Basiswissen wiederholt, konzentriert sich eine systematische Übersicht auf eine eng umrissene und genau definierte Fragestellung. Diese ist häufig genauso formuliert wie das Thema einer wissenschaftlichen Originalarbeit.

Ein Beispiel könnte sein:

„Führt die Verwendung einer Zentrifugalpumpe im Vergleich zur Perfusion mittels einer Rollerpumpe zu weniger zerebrovaskulären Ereignissen im postoperativen Verlauf nach On-pump-CABG?“

Oder:

„Wertigkeit der perioperativen isovolämischen Hämodilution zur Fremdbluteinsparung im Rahmen kardiochirurgischer Eingriffe.“

### **Zielkriterium**

Der Erfolg einer Intervention wird dabei am Anteil der Patienten gemessen, die nach Behandlung in der Interventions- bzw. Kontrollgruppe zerebrovaskuläre Ereignisse aufweisen. Diese Inzidenz stellt somit die gemeinsame Zielgröße aller weiteren Analysen dar. Im Rahmen der zweiten genannten Thematik besteht das Hauptzielkriterium in der Wahrscheinlichkeit (Inzidenz) und der Anzahl der erforderlichen Fremdblutgaben in der Interventions- und Kontrollgruppe. Für diese Zielkriterien müssen definierte Zeitintervalle angegeben werden, nach denen das Auftreten der beschriebenen Endpunkte erfasst werden soll.

Das betrachtete Zeitintervall, innerhalb dessen die Effektivität einer Intervention beurteilt wird, hängt verständlicherweise sehr von der untersuchten klinischen Fragestellung ab.

### **Vollständige Literatursuche**

Der nächste Schritt besteht in einer umfassenden systematischen Literaturrecherche, um möglichst alle Literatur zu dem jeweiligen Thema aufzuspüren. Die Mehrzahl der systematischen Übersichten beschränkt sich auf randomisierte, kontrollierte Studien (Randomised Controlled Trials = RCT). Dass eine systematische Übersicht auch

dann wertvolle Hinweise liefern kann, wenn sie sich auf andere – erst einmal mit geringerer Evidenz belegte – Publikationsformen (z. B. Fallberichte) konzentriert bzw. diese mitberücksichtigt, zeigen insbesondere die Analysen zu seltenen Ereignissen in der Medizin im Allgemeinen, die mit dem klassischen Goldstandard der kontrollierten randomisierten Studie schwer zu erfassen sind [9].

Der Anspruch auf „Vollständigkeit“ ist kein Selbstzweck, sondern dient dazu, den sogenannten Publikationsfehler (publication bias) zu minimieren. „Positive“ oder sogar sensationelle Studienergebnisse werden eher und auch schneller publiziert als „negative“ Studien (also Studienergebnisse, die keine Überlegenheit der untersuchten Intervention demonstrieren), selbst wenn diese die gleiche wissenschaftliche Qualität aufweisen [10]. „Negative“ Resultate werden dagegen vermehrt in kleineren, häufig landessprachlichen Journalen veröffentlicht oder verschwinden nicht selten unpubliziert in den Schreibtischschubladen der Forscher. Aus wirtschaftlichen Interessen kann aber auch ein Sponsor daran interessiert sein, Arbeiten mit „negativem“ Effekt zurückzuhalten, oder es können durch nicht gekennzeichnete Doppelpublikation Effekte einer Intervention suggeriert werden, die so nicht zutreffen [11].

Das Identifizieren von nicht im Format einer Originalie publizierten Daten – der sogenannten „grauen Literatur“ (grey literature) – gelingt nur selten, da eine zentrale Registrierung aller begonnenen klinischen Studien bislang noch in den Kinderschuhen steckt. Man kann aber davon ausgehen, dass zumindest gegenwärtig noch durch die vorrangige Berücksichtigung publizierter Arbeiten in großen, viel gelesenen und leicht verfügbaren Journalen Therapieeffekte tendenziell überschätzt werden. Umso wichtiger ist es daher, durch eine umfassende und gut geplante Literaturrecherche alle verfügbaren Daten, auch die aus schwer zugänglichen Quellen, bei der Metaanalyse zu berücksichtigen. Dabei ist jede Beschränkung, z. B. nur auf englischsprachige Publikationen, erst einmal als inakzeptabel zu betrachten [12], was wiederum nicht bedeutet, dass die Ergebnisse der Metaanalyse für jede Art der medizinischen Fragestellung in Abhängigkeit von der berücksichtigten Spracheinschränkung differieren.

#### „Graue Literatur“

Wenngleich der relative Einfluss der sogenannten grauen Literatur (z. B. Abstracts) auf die Ergebnisse einer Metaanalyse gegenwärtig noch unklar ist – und das Ergebnis

auf diese Frage immer sehr von der betrachteten klinischen Fragestellung und der verfügbaren Literatur abhängen wird –, scheint der Versuch, umfassende Daten aufzufinden und einzuschließen doch mehrheitlich ratsam [13]. Ein gangbarer, wenn auch mühsamer Ausweg, wenigstens an einen Teil der grauen Literatur zu kommen, ist das Durchsuchen von Abstraktbänden von Hand. Auf Kongressen werden häufig Daten präsentiert, die später nie in Form einer Originalarbeit erscheinen. So wurde unlängst das „Schicksal“ von Beiträgen zu einem großen europäischen Anästhesiekongress während der darauffolgenden fünf Jahre verfolgt [14]. Lediglich 40 % aller klinischen Studien erschienen in den darauffolgenden fünf Jahren als Originalarbeit in einem biomedizinischen Journal.

Ein etablierter Algorithmus besteht in der Suche in den drei großen medizinischen Datenbanken Cochrane Controlled Trials Register (Cochrane library), MEDLINE und EMBASE. Neben der Suche nach vorgegebenen MeSH-Begriffen (Medical Subject Headings), ist vielfach eine Freitextsuche mit einer logischen Kombination von Suchbegriffen empfehlenswert. Vor allem in der Cochrane-Bibliothek sind zahlreiche randomisierte kontrollierte Studien als solche gelistet, die in MEDLINE oder EMBASE nicht unter dieser Rubrik verzeichnet sind. Diese beiden Datenbanken weisen wiederum nur eine gemeinsame Deckung von etwa 80 % auf, wobei MEDLINE vorwiegend den amerikanischen Raum abdeckt, während europäische Journale eher in EMBASE vertreten sind.

Eine weitere Möglichkeit besteht in der Kontaktierung der Hersteller pharmazeutischer oder medizintechnischer Produkte, verbunden mit der Bitte, Informationen, insbesondere über unveröffentlichte Untersuchungen, zur Verfügung zu stellen.

Eine intensive Suche nach Daten fördert, wie auch im vorliegenden Fall, häufig Mehrfachpublikationen zutage, die an und für sich ja nicht verwerflich sind, sofern grundlegend andere Aspekte beleuchtet werden und die Daten entsprechend gekennzeichnet sind. Im Einzelfall kann es sehr schwierig zu erkennen sein, ob Daten einer Originalarbeit bereits in Abstraktform veröffentlicht wurden, insbesondere dann, wenn sich die Gesamtpatientenzahl nach der Veröffentlichung von vorläufigen Daten bis zur Publikation der Originalie noch ändert oder die Autoren bzw. deren Reihenfolge variieren. Bleibt dies unerkannt und werden die Daten dann mehrfach in der Metaanalyse verwendet, führt dies

zwangsläufig zu Verzerrungen der wahren Datenlage. Aus diesem Grund lehnen es manche Autoren ab, Daten, die nur in Abstraktform vorliegen, in einer Metaanalyse zu berücksichtigen. Allerdings besteht diese Gefahr auch dann, wenn ausschließlich Originalien berücksichtigt werden [11]. In Originalarbeiten sollte daher vermerkt werden, wenn Daten oder Teile davon bereits in einem Kongressband publiziert wurden. Hilfreich in diesem Zusammenhang wäre auch eine eindeutige Kennzeichnung von Abstracts bzw. Originalarbeiten und eine entsprechende Zitatangabe, um eine Transparenz für die Leser zu gewährleisten.

#### Auswahlprozess der identifizierten Literatur

Im Rahmen der Literatursuche fallen viele Literaturstellen an, die sich bei einer näheren Überprüfung als für die entsprechende Fragestellung nicht relevant herausstellen. Die Selektion, bei welchen Publikationen es sich nun um „relevante Informationen“ handelt, ist ein weiterer entscheidender Schritt bei der Durchführung einer systematischen Übersicht. Die entsprechenden Auswahlkriterien müssen vorab festgelegt sein und der gesamte Selektionsprozess muss in allen Schritten nachvollziehbar sein. In der Praxis haben sich bei Therapiestudien, in denen es um die Bewertung einer Intervention geht, folgende Kriterien durchgesetzt: Beschränkung auf randomisierte, ggf. verblindete Studien, Ausschluss von Doppelpublikationen und Studien ohne inaktive Kontrolle bzw. Placebogruppe. Details über die einzelnen Ausschlussgründe sollten in einem Flussdiagramm transparent dargestellt werden. Eine solche Darstellung wird auch vom bereits erwähnten QUORUM-Statement gefordert [6, 8] und dient in erster Linie dazu, Ein- und Ausschluss bestimmter Artikel für den Leser nachvollziehbar zu gestalten. Der Leser sollte aber den Ausschluss von potenziell relevanter Literatur genau (bzw. die Gründe hierfür) beachten und die klinische Nachvollziehbarkeit für sich überprüfen, da je nach Ausmaß der durchgeführten Einschränkungen erhebliche Variabilität der Ergebnisse zu erwarten ist.

Zusammenfassend dienen definierte Ein- und Ausschlusskriterien dazu, den oben erwähnten Publication-Bias weiter zu minimieren, indem potenziell nicht valide Daten unberücksichtigt bleiben. Dass hierbei gelegentlich die eine oder andere Studie unberechtigterweise „unter den Tisch fällt“, ist weniger der Arroganz der Autoren als vielmehr dem Befolgen von vorab definierten Kriterien anzulasten.

Wurde in der Untersuchung eine adäquate Randomisierung vorgenommen?

Dies ist keine Selbstverständlichkeit, wie das Beispiel einer systematischen Übersicht zeigt, in der zahlreiche klinische Studien aus diesem Grund nicht in die quantitative Analyse eingeschlossen werden konnten, weil Randomisierung beispielsweise zum Zeitpunkt der Durchführung der Studien noch nicht zum etablierten Standard für Therapiestudien gehörte [15]. Die Forderung nach einer adäquaten Randomisierung wird damit begründet, dass Studien, die keine oder keine geeignete Randomisierung aufweisen, den Effekt einer Intervention regelhaft erheblich überschätzen, wie dies beispielhaft für die TENS-Therapie (transkutane elektrische Nervenstimulation) gezeigt werden konnte [16]. Auch Studien, in denen eine Pseudorandomisierung durchgeführt wurde, werden häufig nicht berücksichtigt. Darunter versteht man alle Zuteilungsverfahren, in denen die Untersucher schon vor der Rekrutierung eines Patienten wissen, welcher Gruppe dieser zugehört wird (fehlendes „allocation concealment“). Beispiele hierfür sind Zuteilung im alternierenden Verfahren (ein Patient in der Verum-Gruppe, der nächste in der Kontrollgruppe usw.) oder nach Wochentagen (Montag: Behandlung „A“, Dienstag: Behandlung „B“ usw.).

Handelt es sich um eine prospektive Untersuchung?

Retrospektive Datenerhebungen sowie Fallberichte werden vielfach nicht berücksichtigt, da keine Randomisierung der Patienten stattgefunden haben konnte. Insofern schließt eine Beschränkung auf randomisierte Studien automatisch Publikationen von per definitionem niedrigerer Evidenz aus und ist quasi eine Redundanz. Wie bereits erwähnt, ist diese Beschränkung jedoch nicht apodiktisch zu sehen, da systematische Übersichten, basierend auf anderen Publikationsformen, durchaus von Interesse sein können [9, 17].

Liegt eine Kontrollgruppe vor?

Es werden bei Untersuchungen zu Therapieeffekten in der Regel nur Studien mit einer Kontrollgruppe in die Analyse aufgenommen. Diese ist im Idealfall verblindet [18], doch werden mitunter auch Studien berücksichtigt, die dieses Kriterium nicht erfüllen (z. B. Studien mit einer unbehandelten Kontrollgruppe, vorausgesetzt die Gruppenzuteilung erfolgte randomisiert). Hier besteht in gewissem Maße eine Freiheit der Autoren, sofern die Entscheidung

Beurteilungskriterium	Abstufungen	Punkte
Randomisierung	keine oder inadäquat	0
	adäquat	1
	adäquat + genaue Beschreibung (z. B. verschlossene Umschläge etc.)	2
Verblindung	keine	0
	gegeben	1
	gegeben + genau beschrieben (z. B. kodierte Ampullen etc.)	2
Dropouts (Patienten, deren Daten nicht ausgewertet wurden)	nicht berichtet	0
	berichtet und Gründe angegeben	1

Tab. 1: Oxford-Skala von Jadad, nach der häufig die Qualität der in systematische Übersichten eingeschlossenen Studien beurteilt wird.

begründet ist. Fragestellungen, die zwei aktive Interventionen untersuchen (z. B. Fremdbluteinsparung durch Applikation von Tranexamsäure [Cyclocapron] versus Aprotinin in der Kardiochirurgie) sind jedoch durchaus denkbar und können dazu dienen, die in zwei Metaanalysen über placebokontrollierte Studien gefundene relative Effektivität (jeweils im Vergleich zu Placebo) zu bestätigen bzw. das Effektmaß zu verifizieren. Für die zur Abschätzung der klinischen Bedeutsamkeit einer Intervention so wichtige Quantifizierung der Begleiterscheinungen ist die Kontrolle gegenüber einer Placebo- oder inaktiven Gruppe unerlässlich, zumal wenn kein etablierter „Goldstandard“ existiert.

Handelt es sich um eine Doppelpublikation?

Wie bereits weiter oben angesprochen, dürfen doppelt veröffentlichte Daten auf keinen Fall ein weiteres Mal in die Analyse aufgenommen werden, da sie zu Verfälschung der Therapieeffekte führen können [11].

Wie groß sind die untersuchten Patientengruppen und gibt es geografische bzw. Spracheinflüsse?

Kleine Studien scheinen den Behandlungseffekt tendenziell zu überschätzen [19], weshalb vielfach eine willkürlich gesetzte Grenze (z. B. eine Minimalzahl von zehn Patienten pro Gruppe) als Einschlusskriterium gesetzt wird [20]. Gegebenenfalls erscheint es ratsam, unterschiedliche Subgruppen in puncto geografischer Herkunft und Autorengruppe bzw. Studiengröße zu analysieren und zu prüfen, inwieweit die Ergebnisse hiervon berührt werden und ob die Ergebnisse verallgemeinert werden können [21, 22].

Bewertung der Originalliteratur

Mitunter wird versucht, die Qualität der identifizierten Untersuchungen zu bewer-

ten. Dabei hat sich das in Oxford entwickelte vereinfachte Modell von Jadad [23] durchgesetzt (Tab. 1). Obwohl eine solche Bewertung eigentlich auf die Qualität der Studienplanung und -durchführung abzielt, wird tatsächlich aber eher die Qualität der Publikation, also das Reporting der Studie, bewertet. Das heißt, dass eine unsauber dargestellte Methodik nicht zwangsläufig bedeutet, dass auch die Studie unsauber durchgeführt wurde, und umgekehrt, so dass der Nutzen eines solchen Bewertungsverfahrens umstritten bleibt. Die Beschränkung im zugebilligten Umfang für den Bericht einer Studie, z. B. wenn Ergebnisse als „kurze wissenschaftliche Mitteilung“ publiziert werden, kann auch für den Verzicht der einen oder anderen Angabe verantwortlich sein, die Eingang in die Qualitätsbeurteilung findet.

## DATENEXTRAKTION UND STATISTISCHE ANALYSE (METAANALYSE)

Zur Vorbereitung auf die eigentliche statistische Analyse werden die definierten Zielgrößen aus den Publikationen extrahiert. Im einfachsten Fall sind die Daten dichotom, d. h., das Ereignis (z. B. Transfusion mit Fremdblut bzw. das Auftreten zerebrovaskulärer Ereignisse) ist vorhanden oder nicht vorhanden. Metaanalysen sind aber nicht auf diese Art der Datenpräsentation beschränkt. Vielmehr können auch kontinuierliche Variablen (Mittelwert und Standardabweichung), aber auch Wahrscheinlichkeiten metaanalytisch ausgewertet werden. Obwohl der Begriff „Metaanalyse“ häufig synonym mit „Systematische Übersicht“ verwendet wird, bedeutet er streng genommen nur die statistische Synthese der aus verschiedenen Studien extrahierten Daten. Eine systematische Übersicht, die eine Metaanalyse beinhaltet, wird auch als quantitative systematische Übersicht bezeichnet. Systematische Literaturbetrachtungen ohne statistisches Zusam-

menfassen der Studienergebnisse zu einem Thema werden dagegen als systematische Übersicht (qualitative systematic review) bezeichnet.

*Zusammenfassung ist nicht gleich Addition!*

Wichtig für das Verständnis der statistischen Vorgehensweise bei der Metaanalyse von Studiendaten ist, dass nicht etwa einfach die Zahlen der Patienten mit bzw. ohne Endpunkt zusammenaddiert werden dürfen. Stattdessen werden die Ergebnisse jeder Studie zunächst einzeln berechnet (z. B. das relative Risiko und dessen Konfidenzintervall). Diese Zwischenergebnisse werden dann anhand eines Gewichtungsfaktors zusammengefasst („gepoolt“). Dieser „Umweg“ bei der Berechnung des Gesamteffekts ist erforderlich, da es bei bestimmten Zahlenkonstellationen zu „paradoxen“ Ergebnissen kommen kann, ein Phänomen, das als „Simpson's Paradoxon“ bezeichnet wird. Anhand eines kleinen Zahlenbeispiels soll das verdeutlicht werden (Tab. 2 a bis d).

Der erwähnte Gewichtungsfaktor ergibt sich aus der „Vertrauenswürdigkeit“ der jeweiligen Studienergebnisse. Diese wird wiederum von der untersuchten Patientenzahl und ihrer gleichmäßigen Verteilung auf beide Gruppen bestimmt. Je mehr Patienten untersucht wurden, desto kleiner ist das sich ergebende Konfidenzintervall und desto gewichtiger damit die betreffende Untersuchung. Andererseits wird aber

auch das Ergebnis der Studie im Vergleich zu den Ergebnissen der anderen Untersuchungen gewürdigt. Außenseiterergebnisse werden weniger stark gewichtet als Ergebnisse, die mit der Gesamtheit der anderen Ergebnisse konform sind.

Die zugrundeliegenden mathematischen Modelle sind in einigen Statistikprogrammen enthalten. Für die Durchführung von Metaanalysen gibt es aber auch spezielle Software. Das Programm RevMan der Cochrane-Collaboration steht beispielsweise als Freeware samt umfangreicher Dokumentation, unter anderem zu den verwendeten statistischen Techniken, kostenlos zum Herunterladen aus dem Internet zur Verfügung (<http://www.cc-ims.net/RevMan/RevMan5/download>).

Schließlich muss noch festgelegt werden, welches Modell für die statistische Berechnung verwendet werden soll. Dabei wird überprüft (z. B. mit einer speziellen Modifikation des  $\chi^2$ -Tests), welche der folgenden beiden Annahmen zutrifft:

Die Ergebnisse der eingeschlossenen Studien sind homogen. Das bedeutet, dass die untersuchte Intervention in allen Studien bei mehr oder weniger identischen Ausgangsbedingungen (z. B. Inzidenz zerebrovaskulärer Ereignisse oder Bedarf an Blutprodukten in den Kontrollgruppen) eine vergleichbare Wirkung zeigt. Unter diesen Voraussetzungen kann argumentiert werden, dass einzig und alleine die Intervention (z. B. Applikation von Aprotinin

oder Tranexamsäure) für den beobachteten Effekt verantwortlich ist. In diesem Fall wählt man dann ein Modell der sogenannten festen Effekte (fixed effects model).

Sind im anderen Fall die Ergebnisse dagegen heterogen, d. h., es bestehen wesentliche statistische Unterschiede zwischen den Ergebnissen der einzelnen Studien, wählt man das konservative Modell der zufälligen Effekte (random effects model). Man geht dann davon aus, dass neben der untersuchten Intervention auch andere Ursachen für die beobachteten Ergebnisse mit eine Rolle spielen. Dies können z. B. unterschiedliche Begleitmaßnahmen oder Rahmenbedingungen, verschiedene Operationen oder ein unterschiedliches Patientenkollektiv in den einzelnen Studien sein. Man kalkuliert daher vorsichtiger, um diese Heterogenität zwischen den einzelnen Studien zu berücksichtigen. Jedoch sollte immer geprüft werden, ob die Heterogenität der Studienbedingungen in diesen Fällen nicht zu groß ist, um bei Zusammenfassung der Effekte noch von klinisch relevanten Aussagen zu sprechen.

**ERGEBNISPRÄSENTATION**

Ein Ergebnis im eigentlichen Sinn wird in einem klassischen Übersichtsartikel nur selten präsentiert. Ziel ist dabei vielmehr, Grundlagenwissen zu vermitteln und den Leser mit dem behandelten Gebiet vertraut zu machen. Häufig finden sich nur eher allgemein gehaltene Empfehlungen [24–26].

Da im Rahmen einer systematischen Übersicht eine konkrete Frage gestellt wird, fällt die Antwort dazu wesentlich eindeutiger aus.

Im Folgenden sollen die bereits im Text an einigen Stellen erwähnten Maßzahlen noch einmal systematisch zusammengefasst und näher erläutert werden. Alle drei häufig benutzten Zahlen leiten sich aus einer Vierfeldertafel ab (Tab. 3).

	Gesund	Krank	Inzidenz der Kranken	Relatives Risiko
A	18	2	10 %	1,00
B	72	8	10 %	

Tab. 2 a: Studie 1

	Gesund	Krank	Inzidenz der Kranken	Relatives Risiko
A	50	25	33 %	1,00
B	10	5	33 %	

Tab. 2 b: Studie 2

	Gesund	Krank	Inzidenz der Kranken	Relatives Risiko
Summe A	68	27	28,4 %	2,08
Summe B	82	13	13,7 %	

Tab. 2 c: Studie 1 und 2 einfach addiert

(n)	Kranke (n) / Gesamtzahl (n)		Gewichtung (%)	Relatives Risiko (95 % CI)
	Intervention A (%)	Intervention B (%)		
Studie 1	2 / 20 (10 %)	8 / 80 (10 %)	27,7	1,00 (0,23–4,35)
Studie 2	25 / 75 (30 %)	5 / 15 (30 %)	72,3	1,00 (0,46–2,19)
Studie 1 & 2	27 / 95	13 / 95	100,0	1,00 (0,50–2,01)

Tab. 2 d: Studie 1 und 2 durch gewichtetes Zusammenfassen der relativen Risiken addiert

Tab. 2 a bis d: In diesem Beispiel führt das Addieren der Ergebnisse zweier Einzelstudien, die offensichtlich die identische Wirksamkeit von A und B zeigen, bedingt durch die höhere Inzidenz und größere Patientenzahl in Gruppe A der Studie 2 paradoxerweise zu einem relativen Risiko von 2,08. Dies suggeriert zunächst, Medikament A wäre mit einem mehr als doppelt so großen Risiko für die Erkrankung assoziiert. Die gewichtete Addition führt zum richtigen Ergebnis: Beide Medikamente sind gleich zu bewerten (relatives Risiko: 1,0; 95%-Konfidenzintervall (0,50–2,01).

*Odds-ratio und relatives Risiko*

Odds-ratio (OR) und relatives Risiko (RR) sind in ihrer Interpretation vergleichbar. Sie liefern fast identische Werte, wenn das untersuchte Ereignis eher selten ist oder die Ergebnisse zwischen den Studien sehr ähnlich sind. Die Odds-ratio hat den Vorteil, auch für Fall-Kontroll-Studien bestimmbar zu sein. Für prospektive Studien erscheint dagegen die Angabe des relativen Risikos vorteilhafter [27]. Beide Größen beleuchten die Relation zweier Häufigkeiten. Durch die alleinige Angabe des RR lassen sich aber auch unseriöse Zahlenmanipulationen vornehmen. So muss z. B. ein Medikament, das mit einem RR von 0,5 im

	Zielkriterium (z. B. PÜ&E)	1-Ziel-Kriterium (z. B. kein PÜ&E)	Summe
Intervention	a	b	a+b
Kontrolle	c	d	c+d
Summe	a+c	b+d	a+b+c+d

Tab. 3: Kontingenztafel zur beispielhaften Berechnung der Kenngrößen.

Das zuschreibbare Risiko (Odds-ratio = OR) berechnet sich folgendermaßen:  $OR = (a/b)/(c/d)$ . Nach mathematischer Umformung  $(a * d) / (b * c)$  spricht man auch vom „Kreuzprodukt“ einer Vierfeldertafel. Das relative Risiko (RR) ergibt sich aus dem Quotienten der relativen Häufigkeiten beider Gruppen:  $RR = [a/(a+b)]/[c/(c+d)]$ . Die Number Needed to Treat (NNT) ist der Kehrwert der absoluten Risikoreduktion (ARR), das heißt der Differenz der beiden Gruppenhäufigkeiten:  $ARR = a/(a+b) - c/(c+d)$  •  $NNT = 1/ARR = 1/[a/(a+b) - c/(c+d)]$

merhin die Inzidenz eines unerwünschten Symptoms (z. B. Nachblutung oder Bedarf an einem oder mehr Erythrozytenkonzentraten) halbiert, nicht zwangsläufig auch eine klinisch relevante Wirkung besitzen. Konnte etwa diese Inzidenz nur von z. B. 0,1 % auf 0,05 % verringert werden, ist dies aus klinischer Sicht wohl kaum von großer Bedeutung und rechtfertigt unter Umständen auch den erhöhten Aufwand und die assoziierten Risiken nicht.

#### Auch der absolute Therapieeffekt ist von Bedeutung – absolute Risikoreduktion und Number Needed to Treat

Aus oben Gesagtem geht hervor, dass es erforderlich ist, neben der relativen Risikoreduktion auch die Größe des absoluten Therapieeffektes (absolute Risikoreduktion, ARR) anzugeben. Üblicher als die absolute Risikoreduktion ist der Kehrwert dieser Größe, die Number Needed to Treat (NNT). Die NNT ist mittlerweile eine populäre und akzeptierte Maßzahl [28]. Sie verdeutlicht, wie viele Patienten mit der effektiveren von zwei Interventionen behandelt werden müssen, damit ein Patient, der ohne diese Intervention noch von dem Symptom betroffen wäre, davor bewahrt werden kann. Im Rahmen einer Kosten-Nutzen-Berechnung muss diese Zahl noch mit den Kosten der Intervention multipliziert werden, um den Betrag zu erhalten, der aufgewendet werden muss, um einen Patienten zusätzlich symptomfrei zu halten.

#### Number Needed to Harm

Untersucht man in einer Metaanalyse die Nebenwirkungen einer bestimmten Intervention, spricht man auch von der Number Needed to Harm (NNH). Die NNH repräsentiert, analog zur NNT, die Anzahl an Patienten, unter denen sich genau ein Patient befindet, der bedingt durch die Behandlung an dem unerwünschten Ereignis leidet.

#### Die Basisinzidenz in der Kontroll- oder Placebogruppe ist von Bedeutung

Das Basisrisiko in der Kontroll- oder Placebogruppe wird in quantitativen syste-

matischen Übersichten häufig als Control Event Rate (CER) bezeichnet, das Risiko in der Behandlungsgruppe dagegen als Event Rate (ER). Die CER gibt indirekt Auskunft über die Risikokonstellation des untersuchten Patientenkollektivs. Da die NNT in erheblichem Maße von der CER abhängt, ist ein Vergleich von NNT zwischen Metaanalysen nur unter der Voraussetzung einer vergleichbaren CER (Basisrisiko) zulässig. Es sollten in quantitativen systematischen Übersichtsarbeiten also stets eine Maßzahl für die relative Risikoreduktion (OR oder besser RR) als auch eine Größe für das Ausmaß der absoluten Risikoreduktion (als Maß für die klinische Relevanz des Therapieeffektes) angegeben werden. Da die NNT-Werte stark vom Ausgangsrisiko (CER) abhängen, werden teilweise Studien mit einer definierten Bandbreite des Basisrisikos getrennt von den übrigen ausgewertet. Dadurch gelingt es, die Effektivität einer Intervention für ein bestimmtes Ausgangsrisiko zu bestimmen und der Effektivitätsvergleich zwischen zwei Interventionen, die jeweils in separaten Metaanalysen gegenüber Placebo untersucht wurden, wird indirekt auch über die NNT möglich.

#### GRAFISCHE DARSTELLUNG

Die grafische Darstellung im Rahmen von Metaanalysen erfolgt vielfach unter Zuhilfenahme von charakteristischen Abbildungen.

#### L'Abbé-Plots

Eine erste Abschätzung des Therapieerfolges kann mit Hilfe eines L'Abbé-Plots erfolgen. Dabei werden für jede verfügbare Untersuchung auf der Ordinate und der Abszisse die Häufigkeiten des Zielkriteriums der Behandlungs- bzw. der Kontrollgruppe aufgetragen. Somit repräsentiert jeder Punkt in dem X-Y-Diagramm das Ergebnis einer Studie. Über die Größe des Punktes wird gelegentlich zusätzlich die Gruppengröße der entsprechenden Studie kenntlich gemacht.

Die 45-Grad-Winkelhalbierende repräsentiert die Äquipotenzlinie. Auf ihr liegen alle Studien, bei denen die Inzidenz des

Zielkriteriums in der Verum-Gruppe genauso hoch ist wie in der Kontrollgruppe. Abweichungen nach oben repräsentieren in der Regel Studien, in denen die Inzidenz der Verum-Gruppe niedriger war als die der Kontrollgruppe. Der Gruppenunterschied ist dabei dann umso größer, je weiter die Punkte von der Winkelhalbierenden entfernt liegen (s. Abb. 1). Bereits ohne weitere quantifizierende Analyse kann man in der Beispielabbildung auf einen Blick erkennen, wie ausgeprägt die Therapieeffekte sind. Gegenüber den nachfolgend diskutierten Forest Plots hat der L'Abbé-Plot den Vorteil, den Effekt im Vergleich zu der zugrundeliegenden Placeboinzidenz besonders gut zu visualisieren, was unter Umständen hilft, Außenseiterstudien zu identifizieren.

#### Forest Plots

Die eigentlich charakteristischen Abbildungen im Rahmen von Metaanalysen sind sogenannte Forest Plots. Eine solche Abbildung wurde zum Logo der Cochrane-Collaboration, die sich in besonderem Maße um die Erstellung systematischer Übersichtsarbeiten auf hohem Niveau bemüht (<http://www.cochrane.de>). Es handelt sich dabei um den Forest Plot einer bekannten systematischen Übersicht, welche die Überlegenheit der Kortikosteroidtherapie zur pränatalen Lungenreifeung sichern konnte.

Die Fehlerbalken um einen Punkt oder ein Kreissymbol als Maß für den Punktschätzer der Effektivität bzw. die „Ausläufer“ eines Diamantsymbols (Raute) repräsentieren das 95%-Vertrauensintervall (s. Abb. 2). Mit dieser Wahrscheinlichkeit (0,95) befindet sich das „wahre relative Risiko“ der betreffenden Studie innerhalb der markierten Spanne. Der zusammengefasste („gepoolte“) Therapieeffekt aller Studien wird dann am unteren Ende der Grafik abgebildet. Da dabei die Ergebnisse aller Studien mit einfließen, reduziert sich die statistische Unsicherheit der gesamten Aussage. Dementsprechend verkleinert sich auch das Vertrauensintervall. Vereinbarungsgemäß wertet man das resultierende Endergebnis als „statistisch signifikant“, wenn das 95%-Vertrauensintervall den Wert von 1,0 nicht mehr mit einschließt. Bei der NNT oder NNH gilt dies, wenn das obere Konfidenzintervall unendlich wird.

Bei der Anordnung der Einzelstudien in einem Forest Plot gibt es keine festen Regeln. Man sortiert entweder nach dem relativen Gewicht der Studie im Gesamtkontext, der Effektgröße oder nach dem Jahr der Publikation. Letztere Möglichkeit wird vor allem dann gewählt, wenn eine „ku-

mulative Metaanalyse“ durchgeführt wird. Dabei werden dann nicht mehr die Einzelergebnisse einer jeden Studie aufgeführt, sondern man gibt chronologisch den sich unter Einschluss der jeweiligen Untersuchung ergebenden Gesamteffekt an. Das bedeutet, dass mit dem Erscheinen einer neuen Studie jedes Mal eine neue Berechnung mit dem bis zu diesem Zeitpunkt zur Verfügung stehenden Datenmaterial durchgeführt wird. Damit lässt sich dann sofort erkennen, ab wann ein statistisch signifikanter Gesamteffekt erkennbar war bzw. wie stabil das Ergebnis unter dem Einfluss neu hinzukommender Ergebnisse ist.

Selbstverständlich können diese Daten (RR und Vertrauensintervall) auch in Form von Tabellen präsentiert werden.

Die Forest Plots erschließen recht gut auf grafische Weise den zugrundeliegenden mathematischen Bewertungsalgorithmus.

### **Probleme und Limitationen von systematischen Übersichten und Metaanalysen**

Trotz der zunehmenden Akzeptanz von systematischen Übersichtsarbeiten und der stark ansteigenden Publikationszahlen werden nach wie vor kritische Stimmen bezüglich der Wertigkeit laut.

„Stationen des Erfolgs“ für systematische Übersichten als nützliches Instrument der Entscheidungsfindung waren beispielsweise die Untersuchungen zur Kortikoidgabe bei drohender Frühgeburtlichkeit [29], die wie bereits erwähnt auch im Logo der Cochrane-Collaboration ihren Niederschlag gefunden hat. Auch die positiven Effekte der Lysetherapie nach einem akuten Myokardinfarkt konnten metaanalytisch bereits gesichert werden [30], noch bevor die Ergebnisse in einer großen randomisierten Studie bestätigt wurden [31].

Ein Tiefpunkt dürfte dagegen die von Metaanalysen-Kritikern immer wieder zitierte Arbeit über den günstigen Effekt einer Magnesiumgabe im Rahmen eines Myokardinfarkts gewesen sein. Hier hatte eine Metaanalyse einiger weniger kleiner Studien einen protektiven Effekt der Therapie ergeben [32], der dann durch eine große prospektive Untersuchung widerlegt wurde [33].

Trotz dieses Rückschlags sind die Ergebnisse systematischer Übersichtsarbeiten zunehmend akzeptiert, zumal auch neue Verfahren entwickelt wurden, um die Ergebnisse von Metaanalysen auf eine solidere Basis zu stellen und Ergebnisse im entsprechenden Kontext einzuordnen. Bei der Diskussion darf man zudem nicht außer Acht lassen, dass Metaanalysen per se keine neuen Untersuchungen oder gar ein

magisches Werkzeug sind, die alleine die „ultimative Wahrheit“ herauszufinden in der Lage wären. Somit können auch die Resultate nur so gut und unverfälscht sein wie die klinischen Studien, auf die sie sich stützen. Dementsprechend kritisch müssen daher die Ergebnisse der verfügbaren Originalarbeiten betrachtet werden. Bestehen Zweifel an der Validität der Daten [22, 34] oder Hinweise auf einen Publikationsbias, müssen die Ergebnisse einer Metaanalyse mit großer Zurückhaltung betrachtet werden. Allzu weitreichende – zwangsweise immer retrospektive – Schlussfolgerungen, die sich auf Subgruppenanalysen gründen, müssen gleichermaßen vorsichtig interpretiert werden wie Effekte, die nur in einigen Studien relevant sind, aber den Gesamteffekt der Analyse determinieren und auf die gesamte Population übertragen werden [35]. In Tabelle 4 finden sich Merkmale, anhand deren sich die Qualität einer systematischen Übersicht abschätzen lässt. Aber auch hier gilt: In erster Linie wird der Bericht und nicht primär die Validität der Analyse gemessen.

### **Auffinden von Metaanalysen**

Sorgfältig durchgeführte und korrekt interpretierte systematische Übersichten geben auf einem hohen Evidenzniveau objektive und klinisch verwertbare Antworten auf konkrete Fragestellungen. Die zunehmende Akzeptanz solcher Arbeiten, verbunden mit den organisatorischen Bemühungen der Cochrane-Collaboration, die Erstellung und Bereitstellung systematischer Arbeiten zu fördern, hat zu einer beträchtlichen Zahl derartiger Übersichten geführt. Die Tendenz ist dabei weiterhin steigend [9].

Ähnlich wie bei der Suche nach kontrollierten klinischen Studien, gibt es auch für systematische Übersichten strategisch günstige Suchprofile, die eine optimale Ausbeute liefern. Für die Cochrane-Datenbank systematischer Übersichtsarbeiten sind lediglich themenbezogene Suchwörter nötig. Die Suche nach systematischen Reviews in der Datenbank MEDLINE (z. B. <http://www.nlm.nih.gov>) erfordert dagegen eine weitere Eingrenzung. Dazu wird beispielhaft der von Hunt und McKibbin vorgeschlagene Suchalgorithmus zur Identifizierung systematischer Übersichtsarbeiten zugrunde gelegt [36], wengleich gegenwärtig zumindest in manchen Datenbanken bereits eine Einschränkung auf den Publikationstypus „Metaanalyse“ oder „Review“ durchgeführt werden kann. Letztere Einschränkung umfasst aber nicht notwendigerweise systematische Übersichten. Dieser Algorithmus trägt

Sind Fragestellungen und Methoden klar beschrieben?
Wurden umfassende Suchmethoden benutzt, um relevante Studien zu identifizieren?
Wurden Arbeiten in mehreren Sprachen berücksichtigt?
Wurden Pharmafirmen kontaktiert und um zusätzliche Studienergebnisse gebeten?
Wurden andere Quellen unpublizierter Literatur benutzt (z. B. Abstraktbände, direkte Kontakte zu Forschern, Internet)?
Wurde gezielt nach Doppelpublikationen gefahndet?
Wurde die Methodik der eingeschlossenen Originalarbeiten bewertet und mitgeteilt?
Wurde nach dem Vorliegen eines dominierenden Zentrums bzw. nach Außenseiterergebnissen gefahndet?
Wurde die Robustheit der Ergebnisse in Abhängigkeit von Ein- und Ausschluss bestimmter Studien analysiert oder ist für das Gesamtergebnis lediglich eine kleine Zahl von Studien verantwortlich?
Wurden klinisch sinnvolle Gruppen analysiert und zusammengefasst?
Folgen die Schlussfolgerungen logisch und nachvollziehbar den aufgezeigten Ergebnissen?

Tab. 4: Kriterien, nach denen die Qualität von systematischen Übersichtsarbeiten beurteilt werden kann (modifiziert nach Sackett et al.). Je mehr der vorliegenden Fragen mit „Ja“ beantwortet werden können, desto vertrauenswürdiger ist demnach die Analyse.

der spezifischen Verschlüsselung systematischer Reviews durch den Index der Datenbank MEDLINE Rechnung. Er besteht aus folgenden Schritten:

1. meta-analysis (pt)
2. meta-anal: (textword)
3. metaanal: (textword)
4. quantitative: review: OR quantitative: overview: (textword)
5. systematic: review: OR systematic: overview: (textword)
6. methodologic: review: OR methodologic: overview: (textword)
7. review (pt) AND medline (textword)
8. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7

Dieser Suchalgorithmus wird dann mit der entsprechenden themenbezogenen Suchstrategie mit einer logischen „UND“-Verknüpfung kombiniert.

Schneller zu lokalisieren sind Systematic Reviews über die Menüleiste „Clinical Queries“ unter <http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml> und die Eingabe der thematischen Suchwörter in dem Suchfeld unter: „Find Systematic Reviews“.

### **FAZIT**

Choi und Jadad beschreiben den „narrative review“ als „impressionistic landscape“,

während der „systematic review“ als „classical master drawing“ bezeichnet wird, der für den Leser eine strukturierte und detaillierte Informationsquelle darstellt [10]. Die narrative Übersicht als Vertiefung, Abrundung und Aktualisierung von Lehrbuchinhalten wird folglich nicht von einer neuen Variante der Übersicht abgelöst. Vielmehr geben systematische Übersichtsarbeiten unter anderem durch die Ergebnisse der statistischen Metaanalyse eine objektive, umfassende und quantitative Antwort auf eine ganz gezielte Fragestellung. Die gemeinsame Wissensbasis beider Typen von Übersichtsarbeiten sind jedoch valide klinische Studien, deren Stellenwert durch die Verbreitung der systematischen Übersichten nicht geschmälert wird. Vielmehr wird durch die Technik der systematischen, umfassenden Literatursuche auch bislang wenig beachteten und zitierten – weil z. B. nur in landessprachlichen Journalen veröffentlichten – Untersuchungsergebnissen die ihnen zustehende Würdigung zuteil.

Mit zunehmender Verbreitung systematischer Übersichten müssen bei der Veröffentlichung von Forschungsergebnissen im Rahmen von Originalarbeiten zwingend gewisse Mindeststandards eingehalten werden, wie sie z. B. im CONSORT-Statement festgehalten sind. Dies erleichtert die reibungslose Übernahme der Daten in systematische Übersichtsarbeiten. Traditionelle Übersichtsarbeiten sollten, wenn immer möglich, Aussagen und Ergebnisse von Metaanalysen heranziehen. Auf diese Weise können Behandlungsstrategien hinsichtlich ihrer Effektivität und ihres Risikos objektiver und durch die Wiedergabe leicht verständlicher Maßzahlen wie relativem Risiko oder NNT auch leserfreundlich beurteilt werden.

#### LITERATUR

[1] Antes G, Bassler D, Galandi D: Systematische Übersichtsarbeiten – ihre Rolle in der Evidenzbasierten Gesundheitsversorgung. *Deutsches Ärzteblatt* 1999; 96(Heft 10): 616–622  
 [2] Greene NM: *Anesthesiology journals*, 1992. *Anesth Analg* 1992; 74: 116–120  
 [3] Götz E, Landauer B, Van Aken H: Continuous Medical Education – CME-Zertifizierung für die A&I. *Anaesthesiol Intensivmed* 2001; 42: 5  
 [4] Lindner UK, Oehm V: Die Magie des Impact-Faktors – Enttarnung eines Phänomens. *Anaesthesist* 1997; 46: 1–2  
 [5] Kranke P, Apfel CC, Papenfuss T et al: An increased body mass index is no risk factor for postoperative nausea and vomiting. A systematic review and results of original data. *Acta Anaesthesiol Scand* 2001; 45: 160–166  
 [6] Moher D, Cook DJ, Eastwood S et al: Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Lancet* 1999; 354: 1896–1900

[7] Begg C, Cho M, Eastwood S et al: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *J Am Med Assoc* 1996; 276: 637–639  
 [8] Tramèr MR: CONSORT, QUORUM, and structured abstracts – new rules for authors, new tools for readers. *Eur J Anaesthesiol* 2001; 18: 1–2  
 [9] Wulf H: Epidural anaesthesia and spinal haematoma. *Can J Anaesth* 1996; 43: 1260–1271  
 [10] Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. *Lancet* 1991; 337(8746): 867–872  
 [11] Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ: Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; 315(7109): 635–640  
 [12] Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G: Language bias in randomised trials published in English and German. *Lancet* 1997; 350: 326–329  
 [13] McAuley L, Pham B, Tugwell P, Moher D: Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; 356: 1228–1231  
 [14] Guasch RG, Castillo J, Cifuentes I: The fate of abstracts presented at 1995 European society of anaesthesiologists congress. *Eur J Anaesthesiol* 2001; 18: Suppl 21, A-2  
 [15] Kranke P, Morin AM, Roewer N, Eberhart LHJ: Dimenhydrinate for prophylaxis of postoperative nausea and vomiting. A meta-analysis of randomised controlled trials. *Acta Anaesthesiol Scand* 2002; 46: 238–244  
 [16] Carroll D, Tramèr M, McQuay H, Nye B, Moore A: Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. *Br J Anaesth* 1996; 77: 798–803  
 [17] Tramèr MR, Moore RA, Reynolds DJM, McQuay HJ: Quantitative estimation of rare events which follow a biological progression: a new model applied to chronic NSAID use. *Pain* 2000; 85: 169–182  
 [18] Ernst E, White AR: Acupuncture for back pain: A meta-analysis of randomised controlled trials. *Arch Int Med* 1998; 158: 2235–2241  
 [19] Moore RA, Tramèr MR, Carroll D, Wiffen PJ, McQuay HJ: Quantitative systematic review of typically applied non-steroidal anti-inflammatory drugs. *BMJ* 1998; 316(7128): 333–338  
 [20] Kjellberg F, Tramèr MR: Pharmacological control of opioid-induced pruritus: a quantitative systematic review of randomized trials. *Eur J Anaesthesiol* 2001; 18: 346–357  
 [21] Vickers A, Goyal N, Harland R, Rees R: Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998; 19: 159–166.  
 [22] Kranke P, Apfel CC, Eberhart LH, Georgieff M, Roewer N: The influence of a dominating centre on a quantitative systematic review of gransetron for preventing postoperative nausea and vomiting. *Acta Anaesthesiol Scand* 2001; 45: 659–670  
 [23] Jadad AR, Moore RA, Carroll D et al: Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17(1): 1–12  
 [24] Bardenheuer HJ, Traut F: Inzidenz und Pathophysiologie der postoperativen Übelkeit

und des Erbrechen. *Anaesth Intensivmed Notfallmed Schmerzther* 1997; 32: 617–619  
 [25] Mayr A, Kerger H: Anatomische und pathophysiologische Grundlagen der postoperativen Übelkeit und des postoperativen Erbrechen. *Anaesthesiol Intensivmed* 1999; 40(4): 202–206  
 [26] Apfel CC, Roewer N: Einflussfaktoren von Übelkeit und Erbrechen nach Narkosen: Fiktionen und Fakten. *Anaesthesist* 2000; 49: 629–642  
 [27] Altman DG, Deeks JJ, Sackett DL: Odds ratios should be avoided when events are common. *BMJ* 1998; 317: 7168–1318.  
 [28] Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P: The number needed to treat: a clinically useful nomogram in its proper context. *BMJ* 1996; 312: 426–429  
 [29] Crowley P, Chalmers I, Keirse MJNC: The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. *British Journal of Obstetrics and Gynaecology* 1990; 97: 11–25  
 [30] Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH: Effect of intravenous streptokinase on acute myocardial infarction. Pooled results from randomized trials. *New England Journal of Medicine* 1982; 307: 1180–1182  
 [31] Rovelli F, DeVita C, Feruglio GA, Lotto A, Selvini A, Tognoni G: Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; 397–402  
 [32] Yusuf S, Teo K, Woods K: Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. *Circulation* 1993; 87: 2043–2046  
 [33] ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS4. A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58.050 patients with suspected acute myocardial infarction. *Lancet* 1995; 345: 669–685  
 [34] Kranke P, Apfel CC, Roewer N: Reported data on gransetron and postoperative nausea and vomiting by Fujii et al. are incredibly nice! *Anesth Analg* 2000; 90(4): 1004–1007  
 [35] Rodgers A, Walker N, Schug S et al: Reduction of postoperative mortality and morbidity with epidural or spinal anaesthesia: results from overview of randomised trials. *BMJ* 2000; 321: 1–12  
 [36] Hunt DL, McKibbin KM: Locating and appraising systematic reviews. *Annals of Internal Medicine* 1997; 126: 532–538

Priv.-Doz. Dr. med. Peter Kranke, MBA  
 Klinik und Poliklinik für Anästhesiologie  
 Universitätsklinikum Würzburg  
 Oberdürrbacher Str. 6  
 97080 Würzburg  
 E-Mail: peter.kranke@t-online.de