

# Systematische Übersichtsarbeiten in der Anästhesie – Erstellung, Bedeutung und Interpretation

*Systematic reviews in anaesthesia – methods, implications and interpretation*

P. Kranke<sup>1</sup> und L. H. J. Eberhart<sup>2</sup>

<sup>1</sup> Klinik für Anaesthesiologie der Universität Würzburg (Direktor: Prof. Dr. N. Roewer)

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

**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 kurzem wird die Lektüre solcher Beiträge und die Bearbeitung darauf abgestimmter Fragenkataloge auch von den 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 auch 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 Journals 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 diese Publikationsart typische Begriffe, Abbildungstypen und Maßzahlen zu erklären. Dies erfolgt an einem konkreten Beispiel, einer aktuellen Metaanalyse zur Effektivität von Dolasetron zur Reduktion von Übelkeit und Erbrechen in der postoperativen Phase, die in derselben Ausgabe dieses Journals vorgestellt wird. 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.

**Summary:** As it is the main purpose of original papers to present new research results to the scientific community, these types of publications are rather limited sources of information if the reader is looking for clearly summarized and systematically provided general information on a specific topic. This gap is filled by review and continuous medical education articles, which are much appreciated by many clinicians as an informative literature that serves to deepen and update textbook knowledge. The value of review articles has lately been acknowledged by the fact that CME credit is now attributed to such contributions and their related questionnaires.

However, there are some limitations to traditional reviews that need to be considered. Although they give a comprehensive overview of a particular topic, they usually do not provide relevant efficacy data with respect to specific interventions. Systematic reviews, to the contrary, summarize all available scientific data on a narrow and focused clinical question and – similar to controlled trials – follow well-defined guidelines. Meta-analyses summarize the findings of a systematic review using statistical methods. As they usually allow a precise and comprehensive quantification of the efficacy of specific interventions, they have meanwhile become a most important tool of evidence-based medicine. An initially overwhelming quantity of literature including foreign language publications can thus be condensed to a clinically relevant statement that might be transferred into daily practice.

The aim of this paper is to guide the reader through a typical systematic review, using a meta-analysis of dolasetron in the prevention of postoperative nausea and vomiting for exemplification. The focus is on the methodology of systematic reviews, starting with the initial clinical question up to the graphical presentation of the results. In addition, limitations of systematic reviews, search strategies for the retrieval of systematic reviews, and assessment criteria will be discussed.

**Schlüsselwörter:** Übersichtsarbeiten – Metaanalyse – Statistiken – Ärztliche Weiterbildung

**Key words:** Review – Meta-Analysis – Statistics – Continuous Medical Education.

## Klinische Anästhesie

### Einleitung

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

Die Flut an neuen Publikationen in der Anästhesie ist wie auf vielen anderen Gebieten der Medizin kaum mehr zu überschauen. 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 Leseaufwandes von 19 Artikeln pro Tag, um auf dem aktuellen Stand der Wissenschaft zu bleiben (2). In unserem Fach dürften die Zahlen nicht wesentlich anders aussehen, wenn weltweit bereits im Jahre 1992 mehr als 50 Anästhesiezeitschriften lokalisiert werden konnten (17).

#### **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 ausliegender Zeitschriften in der Regel wenig selektiert. Oft bleibt es dem Zufall überlassen, bei welcher Arbeit der Leser "hängenbleibt". Bei Originalarbeiten kommt es vorrangig auf die exakte Darstellung der Ergebnisse und der verwendeten Methodik an. Schon allein aus Platzgründen kann die Einleitung und der Diskussionsabschnitt das behandelte Thema und die klinischen Hintergründe nur grob anreißen. 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 täglich 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 Fortbildungskonferenzen ("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 (16).

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

Die Autoren dieser Arbeiten sind in aller Regel Wissenschaftler, die sich schon längere Zeit mit dem entsprechenden Thema auseinandersetzt haben. Somit ist es nahezu unvermeidlich, daß beim Abfassen des Manuskripts – bewußt oder unbewußt – auch eigene Meinungen mit einfließen. Weiterhin besteht die Tendenz, eigene Ergebnisse oder die Publikationen befreundeter Arbeitsgruppen überproportional häufig zu zitieren gemäß dem Motto "Zitierst Du mich, zitier' ich Dich" (29), 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 (24). 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 vorgeschrrieben. 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 (32), die teilweise – ähnlich wie das sogenannte CONSORT Statement für kontrollierte klinische Studien (5) – bereits zur Bedingung für die Annahme von systematischen Übersichten in Anästhesiejournalen avanciert ist (40). Dies soll einerseits gewisse Standards gewährleisten, aber andererseits auch im Zuge einer Vereinheitlichung für bessere Lesbarkeit der klinischen Botschaft in den systemati-

schen Übersichten sorgen. Anders als bei der klassischen Übersichtsarbeit, die sich weitgehend von selbst erschließt, finden sich in der Methodik einer systematischen Übersicht häufig wiederkehrende – auf den ersten Blick möglicherweise abschreckende – methodische Vorgehensweisen, die nicht zwangsläufig selbst erklärend sind. Im folgenden soll deshalb der Entstehungsprozeß einer systematischen Übersicht Schritt für Schritt erläutert werden. Dies geschieht am Beispiel einer konkreten systematischen Übersicht zu Dolasetron als antiemetische Prophylaxe nach Narkosen, die in dieser Ausgabe von Anästhesiologie & Intensivmedizin abgedruckt ist (27).

### 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.

Im Beispiel der angesprochenen Arbeit geht es z.B. um die antiemetische Effektivität von Dolasetron, einem neuen Vertreter der 5-HT<sub>3</sub>-Antagonisten, gegenüber Placebo. Die Fragestellungen, wie sie auch im Einleitungsteil der Dolasetron-Arbeit ausformuliert wurden, sind dementsprechend konkret:

- Führt Dolasetron in den untersuchten Dosierungen zur Reduktion von postoperativem Erbrechen (PE) und postoperativer Übelkeit (PÜ)?
- Besteht eine positive Dosis-Wirkungs-Beziehung?
- Wie ist die antiemetische Effektivität einer oralen im Vergleich zur intravenösen Applikation von Dolasetron?

### Zielkriterium

Der Erfolg der antiemetischen Prophylaxe wird dabei am Anteil der Patienten gemessen, die nach Applikation von Dolasetron bzw. Placebogabe während des postoperativen Beobachtungszeitraums Erbrechen bzw. Übelkeit aufwiesen. Diese Inzidenz stellt somit die gemeinsame Zielgröße aller weiteren Analysen dar.

Da die Häufigkeit von Übelkeit und Erbrechen mit zunehmender Zeitdauer nach der Operation auch ohne Intervention spontan abnimmt, können unterschiedlich lange Beobachtungszeiträume nicht beliebig zusammengefaßt werden. Es hat sich daher bewährt, zwei Beobachtungsperioden zu definieren, die ein frühes (0 - 6 Stunden nach der Operation) und ein zusammenfassendes Intervall (0 - 24 Stunden postoperativ) abdecken. Dieses Vorgehen wurde bereits vorher in zahlreichen ähnlichen Analysen angewandt (13, 28, 43) und auch für die vorliegende Arbeit eingesetzt. Das Zielkriterium, wie auch das betrachtete Zeitintervall innerhalb dessen die Effektivität einer Intervention beurteilt wird, hängt verständlicherweise sehr von der untersuchten klinischen Fragestellung ab. Bei einer anderen Begleiterscheinung nach Anästhesien, dem Muskelzittern in der postoperativen Phase, zeigte sich beispielsweise eine sehr ausgeprägte Zeitabhängigkeit der Symptome

nach Applikation der untersuchten Interventionen, so daß lediglich die Effektivitätsdaten mit genau gleichem Zeitintervall nach Applikation der Studienmedikation zusammengefaßt werden konnten (26).

### 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 beschränken 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 Anästhesie und Medizin im allgemeinen, die mit dem klassischen Goldstandard der kontrollierten randomisierten Studie schwer zu erfassen sind (42, 46).

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 (12). "Negative" Resultate werden dagegen vermehrt in kleineren, häufig landessprachlichen Journals 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 (44).

Das Identifizieren von nicht im Format einer Originalie publizierten Daten – der sogenannten "grauen Literatur" – gelingt nur selten, da bislang keine verbindliche und zentrale Registrierung aller begonnenen klinischen Studien erfolgt. Man kann aber davon ausgehen, daß zumindest gegenwärtig noch durch die vorrangige Berücksichtigung publizierter Arbeiten in großen, viel gelesenen und leicht verfügbaren Journals 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 (14), was wiederum nicht bedeutet, daß die Ergebnisse der Metaanalyse für jede Art der medizinischen Fragestellungen in Abhängigkeit von der berücksichtigten Spracheinschränkung differieren.

### "Graue Literatur"

Wenngleich der relative Einfluß der sogenannten grauen Literatur (z.B. Abstrakts) auf die Ergebnisse

## Klinische Anästhesie

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 (31). Ein gangbarer, wenn auch mühsamer Ausweg, wenigstens an einen Teil der grauen Literatur zu kommen, ist das Durchsuchen von Abstractbänden 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 darauf folgenden fünf Jahre verfolgt (18). Lediglich 40% aller klinischen Studien erschienen in den darauf folgenden fünf Jahren als Originalarbeit in einem biomedizinischen Journal.

Um den Anforderungen nach einer umfassenden Literatursuche gerecht zu werden, wurden im Rahmen der vorliegenden Analyse zu Dolasetron mehrere Verfahren kombiniert. Zum einen erfolgte die 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") wurde auch eine Freitextsuche mit einer logischen Kombination von Suchbegriffen verwendet. 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. Zudem wurde die Herstellerfirma von Dolasetron angeschrieben und gebeten, Informationen, insbesondere über unveröffentlichte Untersuchungen, zur Verfügung zu stellen. Eine manuelle Suche in den Literaturverzeichnissen aller bis dahin identifizierten Arbeiten sowie in den einschlägigen Anästhesiekongreßbänden ab 1994 rundete die Literaturrecherche ab. Das Ergebnis dieser Suchstrategien wird in einem Flußdiagramm (Abbildung 1, siehe Seite 415) der korrespondierenden Dolasetron-Analyse dargestellt. Vor allem die Handsuche erbrachte vier neue, bislang noch nicht publizierte Daten, aber auch schon seit längerem bekannte Daten, die in der Zwischenzeit bereits als Originalarbeiten veröffentlicht worden waren. In zwei Fällen war es erforderlich, die Autoren der Arbeiten direkt zu kontaktieren, um die anfänglich noch unvollständigen Informationen für die quantitative Analyse nutzen zu können.

Eine intensive Suche nach Daten fordert, 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 werden. Im Einzelfall kann es sehr schwierig zu erkennen sein, ob Daten einer Originalarbeit bereits in

Abstractform 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 Abstractform vorliegen, in einer Metaanalyse zu berücksichtigen. Allerdings besteht diese Gefahr auch dann, wenn ausschließlich Originalien berücksichtigt werden (44). 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 folgende Kriterien durchgesetzt: Beschränkung auf randomisierte, ggf. verblindete Studien, Ausschluß von Doppelpublikationen und Studien ohne inaktive Kontrolle bzw. Placebogruppe. Diese wurden auch in der vorliegenden Metaanalyse zum Dolasetron angewendet. Details über die einzelnen Ausschlußgründe in dieser Arbeit finden sich wiederum im Flußdiagramm (Abbildung 1, siehe Seite 415). Eine solche Darstellung wird auch vom bereits erwähnten QUORUM-Statement gefordert und dient in erster Linie dazu, Ein- und Ausschluß bestimmter Artikel für den Leser nachvollziehbar zu gestalten. Der obere Teil des Flußdiagramms (oberhalb des Statements über "Potentiell relevante randomisierte kontrollierte Studien") ist dabei weniger von Relevanz und wird je nach subjektiver Betrachtungsweise der Autoren variieren. Der Leser sollte aber den Ausschluß bzw. die Gründe hierfür von potentiell relevanter Literatur genau 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 potentiell 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.

## Systematische Übersichtsarbeiten

### *Wurde in der Untersuchung eine adäquate Randomisierung vorgenommen?*

Dies war bei allen recherchierten Originalarbeiten zu Dolasetron der Fall, ist aber dennoch keine Selbstverständlichkeit, wie das Beispiel einer anderen 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 (28). 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 gezeigt werden konnte (7). 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 zugehören wird (fehlendes "allocation concealment"). Beispiele hierfür sind Zuteilung im alterierenden Verfahren (ein Patient in der Verumgruppe, 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 können nicht berücksichtigt werden, da keine Randomisierung der Patienten stattgefunden haben konnte. Insofern schließt eine Beschränkung auf randomisierte Studien automatisch Publikationen von per Definition 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 (42, 46).

### *Liegt eine Kontrollgruppe vor?*

Es werden bei Untersuchungen zu Therapieeffekten nur Studien mit einer Kontrollgruppe in die Analyse aufgenommen. Diese ist im Idealfall verblindet (15), doch werden in der Regel 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 begründet ist. Fragestellungen, die zwei aktive Interventionen untersuchen (z.B. Droperidol gegenüber Ondansetron) 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. 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, wie dies auch bei der

Prophylaxe von Übelkeit und Erbrechen in der postoperativen Phase nach wie vor der Fall ist.

### *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 (44).

### *Wie groß sind die untersuchten Patientengruppen, und gibt es geographische bzw. Spracheinflüsse?*

Kleine Studien scheinen den Behandlungseffekt tendenziell zu überschätzen (33), weshalb vielfach eine willkürlich gesetzte Grenze (z.B. eine Minimalzahl von zehn Patienten pro Gruppe) als Einschlußkriterium gesetzt wird (22). Gegebenenfalls erscheint es ratsam, unterschiedliche Subgruppen in puncto geographischer Herkunft und Autorengruppe zu analysieren und zu prüfen, inwieweit die Ergebnisse hiervon berührt werden und ob die Ergebnisse verallgemeinert werden können (23, 45).

### *Bewertung der Originalliteratur*

Mitunter wird versucht, die Qualität der identifizierten Untersuchungen zu bewerten. Dabei hat sich das in Oxford entwickelte, vereinfachte Modell von Jadad (21) durchgesetzt (Tab. 1). Obwohl eine solche Bewertung eigentlich auf die Qualität der Studienplanung und -durchführung abzielt, wird tatsächlich aber eher die Qualität der Publikation 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. Der Vollständigkeit halber wurden aber auch in der korrespondierenden Dolasetron-Analyse alle eingeschlossenen Studien mit dem Oxford-Score bewertet (Tab. 1 der Dolasetron-Analyse).

### **Datenextraktion und statistische Analyse (Metaanalyse)**

Zur Vorbereitung auf die eigentliche statistische Analyse werden die definierten Zielgrößen aus den Publikationen extrahiert. Im vorliegenden Beispiel sind die Daten dichotom, d.h. das Ereignis (z.B. PÜ) 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 systema-

## Klinische Anästhesie

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

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

tische Übersicht bezeichnet. Systematische Literaturbetrachtungen ohne statistisches Zusammenfassen der Studienergebnisse zu einem Thema werden dagegen als systematische Übersicht (systematic review) bezeichnet.

### Zusammenfassung ist nicht gleich Addition!

Wichtig für das Verständnis der statistischen Vorgehensweise bei der Metaanalyse der Daten ist, daß nicht etwa einfach die Zahlen der Patienten mit bzw. ohne PONV zusammen addiert werden dürfen. Statt dessen 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 "paradoxa" Ergebnissen kommen kann, ein Phänomen, das als "Simpson's Paradoxon" bezeichnet wird. Anhand eines kleinen Zahlenbeispiels soll das verdeutlicht werden (Tab. 2a - d).

Der erwähnte Gewichtungsfaktor (er wird in der besprochenen Metaanalyse in Abbildung 4a - 4d (siehe Seite 421 - 422) für jede einzelne Studie aufgeführt) 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 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.cochrane.de/cochrane/dnld-germ.htm#REVM>). Es diente auch für die Analyse der Dolasetron-Daten.

Schließlich muß noch festgelegt werden, welches Modell für die statistische Berechnung verwendet werden soll. Dabei wird überprüft (z.B. mit einer speziellen Modifikation des  $\chi^2$ -Testes), 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. PONV-Inzidenz in den Kontrollgruppen) eine vergleichbare Wirkung zeigt. Unter diesen Voraussetzungen (die im gewählten Beispiel nur zum Teil gegeben sind) kann argumentiert werden, daß einzige und alleine die Intervention (Applikation von Dolasetron) für den beobachteten Effekt verantwortlich ist. In diesem Fall wählt man dann ein Modell der sogenannten festen Effekte ("fixed effect 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 einer Rolle spielen. Im vorliegenden Fall können dies z.B. unterschiedliche Narkoseverfahren, 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 klas-

## Systematische Übersichtsarbeiten

**Tabelle 2a-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)).

**Tabelle 2a: Studie 1:**

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

**Tabelle 2b: Studie 2:**

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

**Tabelle 2c: Studie 1 und 2 einfach addiert:**

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

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

Studien (n)	Kranke (n) / Gesamtzahl (n) Intervention A (%)	Intervention B (%)	Gewichtung (%)	Relatives Risiko (95% CI)
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	<b>1,00 (0,50 - 2,01)</b>

sischen Übersichtsartikel nur selten präsentiert. Ziel ist ja dabei vielmehr, Grundlagenwissen zu vermitteln und den Leser mit dem behandelten Gebiet vertraut zu machen. Häufig finden sich nur eher allgemein gehaltene Empfehlungen, so auch in klassischen Übersichtsarbeiten zum Thema PONV (3, 4, 6, 30).

Da im Rahmen einer systematischen Übersicht eine konkrete Frage gestellt wird, fällt die Antwort dazu wesentlich eindeutiger aus. Danach ist das Risiko, nach Dolasetrongabe postoperativ unter Übelkeit und Erbrechen zu leiden, nur knapp drei Viertel so hoch wie ohne Therapie (gepooltes relatives Risiko = 0,78). Das bedeutet, daß bei einem mittleren Risiko für PÜ&E von 76% ohne antiemetische Prophylaxe etwa sechs Patienten behandelt werden müssen, damit ein Patient, der ohne Dolasetron unter PONV gelitten hätte, nun frei davon bleibt (gepoolte NNT = 5,9).

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).

#### Odds-ratio und relatives Risiko

Odds-ratio (OR) und relatives Risiko (RR) sind in ihrer Interpretation vergleichbar. Sie liefern fast iden-

tische 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 (1). 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 immerhin die Inzidenz eines unerwünschten Symptoms (z. B. PONV) halbiert, nicht zwangsläufig auch eine klinisch relevante Wirkung besitzen. Konnte etwa die Inzidenz von PONV nur von 10% auf 5% verringert werden, ist dies aus klinischer Sicht wohl kaum von großer Bedeutung.

*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 Therapieeffekts (absolute Risikoreduktion) anzugeben. In unserem Beispiel beträgt sie lediglich fünf Prozentpunkte (oder 0,05 als Fraktion von "1"). Wie bereits erwähnt, gibt man üblicherweise den Kehrwert dieser Größe, die NNT, an:  $NNT = 1 / 0,05 = 20$ .

## Klinische Anästhesie

**Tabelle 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:

$$ARR = a / (a+b) - c / (c+d)$$

$$NNT = 1 / ARR = 1 / (a / (a+b)) - (c / (c+d))$$

	Zielkriterium (z.B. PÜ&E)	1-Zielkriterium (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

Die NNT ist mittlerweile eine populäre und akzeptierte Maßzahl (8). 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. Im Fall von Dolasetron liegt die NNT für die prophylaktische Gabe bei sechs (für die 12,5 mg Dosierung). Setzt man für diese Dosis einen Einkaufspreis von 3,6 Euro (7 DM) an, so müssen für einen Patienten, der durch die Prophylaxe mit Dolasetron zusätzlich frei von emetischen Symptomen bleibt, etwa  $3,6 \times 6 = 21,6$  Euro aufgewendet werden. Dies gilt - wie bereits gesagt - aber nur für das jeweils angegebene Basisrisiko (behelfsmäßig abgeschätzt über die Inzidenz in der Placebogruppe).

### 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. Im Fall der Dolasetron-Analyse konnte z.B. gezeigt werden, daß von 50 mit Dolasetron behandelten Patienten einer zusätzlich Kopfschmerzen entwickelt (NNH = 50). Dies war die einzige statistisch zu sichernde Nebenwirkung. Das heißt, für die übrigen aufgetretenen Nebenwirkungen lag die obere Schranke des 95%-Vertrauensintervalls

der NNH im Unendlichen bzw. schloß den Wert 1,0 ein (bei Betrachtung des relativen Risikos).

*Die Basisinzidenz in der Placebogruppe ist von Bedeutung*

Das Basisrisiko in der Placebogruppe wird in quantitativen systematischen 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, z.B. bei Prophylaxestudien zu PÜ&E (41). 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. Das relative Risiko, nach Prophylaxe mit Dolasetron unter PÜ&E zu leiden, beträgt z.B. 0,78 bei einer CER von 76% in den Placebogruppen. Wäre beispielsweise die CER nur halb so hoch, würde die entsprechende NNT von sechs auf Werte um elf hochschnellen. Diese Zahlen Spiele machen somit deutlich, daß sowohl 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 sollte. 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 (z.B. Propofolanästhesie gegenüber Ondansetron-Gabe bei Inhalationsanästhesie), wird indirekt auch über die NNT möglich.

### Graphische Darstellung

Die graphische 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.

In Abbildung 2 und 3 (siehe Seite 420) der begleitenden Dolasetron-Metaanalyse ist eine derartige Graphik wiedergegeben. Die 45-Grad Winkelhalbierende repräsentiert die Äquipotenzlinie. Auf ihr liegen alle Studien, bei denen die Inzidenz des Zielkriteriums in der Dolasetrongruppe genauso hoch ist wie in der Kontrollgruppe. Alle Abweichungen nach oben (im Fall von Dolasetron alle Punkte) repräsentieren Stu-

dien, in denen die PÜ&E-Inzidenz der Dolasetrongruppe niedriger war als die der Kontrollgruppe. Umgekehrt wären Studien, in denen die Placebo-Gruppe besser abschneidet, unterhalb der Winkelhalbierenden aufgetragen. Der Gruppenunterschied ist dabei um so größer, je weiter die Punkte von der Winkelhalbierenden entfernt liegen. Bereits ohne weitere quantifizierende Analyse kann man hier auf einen Blick erkennen, daß sich die Therapieeffekte eher in Grenzen halten. Die Mehrzahl der Punkte liegen nur wenig oberhalb der Winkelhalbierenden. Gegenüber den nachfolgend diskutierten Forestplots 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.

#### *Forestplots*

Die eigentlich charakteristischen Abbildungen im Rahmen von Metaanalysen sind sogenannte Forestplots. 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 Forestplot einer bekannten systematischen Übersicht, welche die Überlegenheit der Kortikosteroidtherapie zur pränatalen Lungenreifung sichern konnte.

Entsprechende Forestplots für die Effektivität von Dolasetron befinden sich in den Abbildungen 4a - c (siehe Seite 421 - 422) der korrespondierenden Arbeit. Dabei ist für jede Studie eine typische Kenngröße, in diesem Fall das relative Risiko, dargestellt. Die Fehlerbalken repräsentieren das 95%-Vertrauensintervall. 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 Graphik 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 Forestplot gibt es keine festen Regeln. Man sortiert entweder wie im vorliegenden Fall 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 "kumulative 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 die Daten auch in Form von Tabellen präsentiert werden, wie dies in der Dolasetronanalyse auch erfolgt ist.

Die Forestplots erschließen recht gut auf graphische Weise den zugrundeliegenden mathematischen Bewertungsalgorithmus. So zeigen sich in der entsprechenden Abbildung 4a (PÜ über 24 Stunden) folgende Effekte: Einerseits nimmt das Gewicht einer Studie mit zunehmender Patientenzahl zu, und eine große (multizentrische) Studie trägt mit ca. 14,1% deutlich mehr zum Gesamteffekt bei als die Studie mit der geringsten Gewichtung (lediglich 0,4%). Zum anderen wird aber bei den zwei kleinsten Studien dieser Abbildung auch deutlich, daß Abweichungen vom Effekt der überwiegenden Studienergebnisse eher mit einem geringeren Gewicht in die gepoolte Analyse eingehen (Rüsch et al. versus Eberhart et al.: Letztere Studie erhielt hier eine geringere Gewichtung, obwohl die Patientenzahl sogar höher war).

#### **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 (11), 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 (36), noch bevor die Ergebnisse in einer großen randomisierten Studie bestätigt wurden (35).

Ein Tiefpunkt dürfte dagegen die von Metaanalysenkritikern 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 (47), der dann durch eine große prospektive Untersuchung widerlegt wurde (20).

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 magische Werkzeuge 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

## Klinische Anästhesie

verfügbarer Originalarbeiten betrachtet werden. Bestehen Zweifel an der Validität der Daten (23, 25) oder Hinweise auf einen Publikationsbias, müssen die Ergebnisse einer Metaanalyse mit großer Zurückhaltung betrachtet werden. Allzu weitreichende – zwangsläufig 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 (34). In Tabelle 4 finden sich Merkmale, anhand derer 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.

Im Fall der hier diskutierten systematischen Übersichtsarbeit zu Dolasetron wurden zunächst kleinere homogene Gruppen getrennt analysiert. Dieses Vorgehen war notwendig, da bereits im Vorfeld bekannt war, daß sehr heterogene Dosierungen und Applikationsformen (von 12,5 mg intravenös bis zu 200 mg oral) zum Teil auch im Rahmen von Dosisfindungsstudien untersucht worden waren. Erst nachdem sich die Ergebnisse dieser Teilanalysen als sehr homogen herausgestellt hatten, wurden Untersuchungen mit unterschiedlichen Dosierungen und Applikationstechniken zusammengeführt und dabei auch die Ergebnisse noch nicht publizierter Daten mit eingeschlossen. Dieses Vorgehen erlaubt einerseits gezielte Aussagen für eine größere Zahl homogener Untergruppen von Studien (Dosierungen und Applikationsformen), andererseits aber auch eine zusammenfassende Bewertung des globalen Therapieeffekts von Dolasetron wie sie vor allem für die klinische Praxis als "Take home message" von Interesse ist.

### 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 (<http://www.cochrane.de/cochrane/newreviews.htm>) 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 McKibbon vorgeschlagene Suchalgorithmus zur Identifizierung systematischer Übersichtsarbeiten zugrunde gelegt (19), wenngleich gegenwärtig zumindest in manchen Datenbanken

**Tabelle 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.

- |   |   |
|---|---|
| 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. Abstractbände, 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?  |   |

bereits eine Einschränkung auf den Publikationstypus "Metaanalyse" durchgeführt werden kann. Letztere Einschränkung umfaßt aber nicht notwendigerweise systematische Übersichten. Dieser Algorithmus trägt 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.

### Metaanalysen im Bereich Anästhesie, Intensivmedizin, Notfallmedizin und zur Schmerztherapie

Auf dem Gebiet der Anästhesie und der benachbarten Gebiete gibt es zahlreiche systematische Übersichtsarbeiten, die von Choi und Mitarbeitern kürzlich zusammengestellt wurden (9). Tramèr gibt ebenfalls einen guten Überblick zu zahlreichen interessanten Arbeiten, wobei die gelungene Kommentierung der jeweiligen Arbeiten hervorzuheben ist (37). Speziell für die Belange der Anästhesisten sind ebenfalls zahlreiche Seiten im Internet verfügbar, von denen zwei willkürlich ausgewählte hier erwähnt sein sollen. Zum einen ist dies die Seite der Oxford Pain Relief Unit, die mit einer speziellen der Schmerztherapie gewidmeten Homepage unter <http://www.jr2.ox.ac.uk/bandolier/booth/painpag/index.html> zu erreichen ist. Unter <http://www.hcuge.ch/anesthesie/anglais/evidence/arevusyst.htm> ist die Seite von Dr. Martin Tramèr in Genf, der einen wesentlichen Beitrag zu den bislang verfügbaren systematischen Übersichten im Bereich der perioperativen Medizin geleistet hat, zu finden. Speziell zum Themenkomplex Übelkeit und Erbrechen nach Narkosen gibt es mittlerweile zahlreiche systematische Übersichtsarbeiten, die bereits heute konkrete und EBM-basierte Therapieempfehlungen zulassen (38, 39).

### 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 Journals 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 objek-

tiver und durch die Wiedergabe leicht verständlicher Maßzahlen wie relativem Risiko oder NNT auch leserfreundlich beurteilt werden.

### Danksagung

Wir danken herzlich Herrn Prof. Dr. med. Clemens-Alexander Greim und Frau Dr. med. Astrid M. Morin, DEAA, für die kritische und konstruktive Durchsicht des Manuskripts.

### Literatur

- Altman DG, Deeks JJ, Sackett DL: Odds ratios should be avoided when events are common. BrMedJ 317 (1998) 7168
- Antes G, Bassler D, Galandi D: Systematische Übersichtsarbeiten - Ihre Rolle in der Evidenz-basierten Gesundheitsversorgung. Deutsches Ärzteblatt 96 (1999) 616
- Apfel CC, Roewer N: Einflußfaktoren von Übelkeit und Erbrechen nach Narkosen: Fiktionen und Fakten. Anaesthesist 49 (2000) 629
- Bardenheuer HJ, Traut F: Inzidenz und Pathophysiologie der postoperativen Übelkeit und des Erbrechens. Anaesthesiol Intensivmed Notfallmed Schmerzther 32 (1997) 617
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. J Am Med Assoc 276 (1996) 637
- Biedler A, Wilhelm W: Postoperative Übelkeit und Erbrechen. Anaesthesist 47 (1998) 145
- 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 77 (1996) 798
- Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P: The number needed to treat: a clinically useful nomogram in its proper context. BrMedJ 312 (1996) 426
- Choi PTL, Halpern SH, Malik N, Jadad AR, Tramèr MR, Walder B: Examining the evidence in anesthesia literature: A critical appraisal of systematic reviews. Anesth Analg 92 (2001) 700
- Choi PTL, Jadad AR: Systematic reviews in anesthesia: should we embrace them or shoot the messenger? Can J Anaesth 47 (2000) 486
- Crowley P, Chalmers I, Keirse MJNC: The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. Br J Obstet Gynaecol 97 (1990) 11
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. Lancet 337 (1991) 867
- Eberhart LHJ, Morin AM, Bothner U, Georgieff M: Droperidol im Vergleich zu 5-HT3-Antagonisten zur Prophylaxe von Übelkeit und Erbrechen in der postoperativen Phase. Anaesthesiol Intensivmed 42 (2001) 58
- Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G: Language bias in randomised trials published in English and German. Lancet 350 (1997) 326
- Ernst E, White AR: Acupuncture for back pain: A meta-analysis of randomised controlled trials. Arch Int Med 158 (1998) 2235
- Götz E, Landauer B, Van Aken H: Continuous Medical Education - CME-Zertifizierung für die A&I. Anaesthesiol Intensivmed 42 (2001) 5
- Greene NM: Anesthesiology journals, 1992. Anesth Analg 74 (1992) 116

## Klinische Anästhesie

18. *Guasch RG, Castillo J, Cifuentes I:* The fate of abstracts presented at 1995 european society of anaesthesiologists congress. Eur J Anaesthesiol 18 (2001) Suppl 21, A-2
19. *Hunt DL, McKibbon KM:* Locating and appraising systematic reviews. Ann Intern Med 126 (1997) 532
20. 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 58050 patients with suspected acute myocardial infarction. Lancet 345 (1995) 669
21. *Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ:* Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 17 (1996) 1
22. *Kjellberg F, Tramèr MR:* Pharmacological control of opioid-induced pruritus: a quantitative systematic review of randomized trials. Eur J Anaesthesiol 18 (2001) 346
23. *Kranke P, Apfel CC, Eberhart LH, Georgieff M, Roewer N:* The influence of a dominating centre on a quantitative systematic review of granisetron for preventing postoperative nausea and vomiting. Acta Anaesthesiol Scand 45 (2001) 659
24. *Kranke P, Apfel CC, Papenfuss T, Rauch S, Loebmann U, Greim C-A, Roewer N:* 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 45 (2001) 160
25. *Kranke P, Apfel CC, Roewer N:* Reported data on granisetron and postoperative nausea and vomiting by Fujii et al. are incredibly nice! Anesth Analg 90 (2000) 1004
26. *Kranke P, Eberhart LH, Roewer N, Tramèr MR:* Pharmacological treatment of postoperative shivering: a quantitative systematic review of randomized controlled trials. Anesth Analg 94 (2002) 453
27. *Kranke P, Eberhart LHJ, Morin AM, Roewer N:* Dolasetron zur Prophylaxe von Übelkeit und Erbrechen nach Narkosen - Eine Metaanalyse kontrollierter randomisierter Studien. Anaesthesiol Intensivmed (2002) 413
28. *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 46 (2002) 238
29. *Lindner UK, Oehm V:* Die Magie des Impact Faktors - Enttarnung eines Phänomens. Anaesthesist 46 (1997) 1
30. *Mayr A, Kerger H:* Anatomische und pathophysiologische Grundlagen der postoperativen Übelkeit und des postoperativen Erbrechens. Anaesthesiol Intensivmed 40 (1999) 202
31. *McAuley L, Pham B, Tugwell P, Moher D:* Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 356 (2000) 1228
32. *Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF:* for the QUORUM Group. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. Lancet 354 (1999) 1896
33. *Moore RA, Carroll D, Wiffen PJ, Tramèr M, McQuay HJ:* Quantitative systematic review of topically-applied non-steroidal anti-inflammatory drugs. BrMedJ 316 (1998) 333
34. *Rodgers A, Walker N, Schug S, McKee A, Kehlet H, van Zundert A, Sage D, Futter M, Saville G, Clark T, MacMahon S:* Reduction of postoperative mortality and morbidity with epidural or spinal anaesthesia: results from overview of randomised trials. BrMedJ 321 (2000) 1
35. *Rovelli F, DeVita C, Feruglio GA, Lotto A, Selvini A, Tognoni G:* Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet (1986) 397
36. *Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH:* Effect of intravenous streptokinase on acute myocardi-
- al infarction. Pooled results from randomized trials. N Engl J Med 307 (1982) 1180
37. *Tramèr MR (Hrsg.):* Evidence based resource in anaesthesia and analgesia. London: BrMedJ Bookshop 2000
38. *Tramèr MR:* A rational approach to the control of post-operative nausea and vomiting: evidence from systematic reviews. Part 1. Efficacy and harm of antiemetic interventions, and methodological issues. Acta Anaesthesiol Scand 45 (2001) 4
39. *Tramèr MR:* A rational approach to the control of post-operative nausea and vomiting: evidence from systematic reviews. Part 2. Recommendations for prevention and treatment, and research agenda. Acta Anaesthesiol Scand 45 (2001) 14
40. *Tramèr MR:* CONSORT, QUORUM, and structured abstracts - new rules for authors, new tools for readers. Eur J Anaesthesiol 18 (2001) 1
41. *Tramèr MR, Moore A, McQuay H:* Propofol anaesthesia and postoperative nausea and vomiting: quantitative systematic review of randomized controlled studies. Br J Anaesth 78 (1997) 247
42. *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 85 (2000) 169
43. *Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ:* Efficacy, dose-response, and safety of ondansetron in prevention of postoperative nausea and vomiting: a quantitative systematic review of randomized placebo-controlled trials. Anesthesiology 87 (1997) 1277
44. *Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ:* Impact of covert duplicate publication on meta-analysis: a case study. BrMedJ 315 (1997) 635
45. *Vickers A, Goyal N, Harland R, Rees R.:* Do certain countries produce only positive results? A systematic review of controlled trials. Control Clin Trials 19 (1998) 159
46. *Wulf H.:* Epidural anaesthesia and spinal haematoma. Can J Anaesth 43 (1996) 1260
47. *Yusuf S, Teo K, Woods K:* Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. Circulation 87 (1993) 2043.

### Korrespondenzadresse:

Dr. med. Peter Kranke  
Klinik für Anaesthesiologie  
Universität Würzburg  
Josef-Schneider-Straße 2  
D-97080 Würzburg.