

# SYSTEMATISCHE ÜBERSICHTSARBEITEN - EINE EINFÜHRUNG IN DIE ZUSAMMENFASSUNG UND BEWERTUNG VON ORIGINALARBEITEN AM BEISPIEL KLINISCHER STUDIEN

KLAUS LINDE, DIETER MELCHART<sup>1</sup>, FLORIAN EITEL<sup>2</sup>, MÜNCHEN

## Zusammenfassung

Angesichts der Flut wissenschaftlicher Studien sind systematische Übersichtsarbeiten von großer Relevanz für Ärzte, Wissenschaftler, politische Entscheidungsträger und Patienten. Übersichtsarbeiten fassen Methoden und Ergebnisse vorliegender Originalarbeiten zusammen. Sie werden dann als systematisch bezeichnet, wenn sie nach einem vorher festgelegten Studienprotokoll erstellt werden, in dem Ziele und methodischen Schritte soweit möglich festgelegt sind. Systematische Übersichtsarbeiten können in fast allen Bereichen empirischer Forschung einschließlich der Ausbildungsforschung Anwendung finden. Das Konzept und die Methoden beinhalten jedoch eine ganze Reihe von Fehlermöglichkeiten, deren bekannteste der Vergleich des Unvergleichbaren ("Äpfel und Birnen") ist. Für eine angemessene Interpretation sind daher Kenntnisse der wichtigsten Vorgehensweisen und Probleme systematischer Übersichtsarbeiten wünschenswert. Im vorliegenden Artikel werden die einzelnen Schritte bei der Erstellung und Bewertung einer systematischen Übersicht im Bereich klinischer Studien dargestellt: Fragestellung, Literatursuche und Studienselektion, Extraktion von Informationen, Qualitätsbewertung, Zusammenfassung und Darstellung der Ergebnisse.

## Summary

Due to the huge amount of new information systematic reviews of the existing scientific evidence on a specific topic are crucial for health care professionals and patients. Reviews summarize methods and results of available original studies; they are called systematic if performed according to a protocol predefining the research question and the review methods. Systematic reviews can be done in most fields of empirical research, including educational research. Inherent to the methods of systematic reviews, there are a number of possible flaws, the most popular being the comparison of non-comparable studies ("apples and oranges"). For an adequate interpretation of systematic reviews, basic knowledge of how such studies are planned, performed and analyzed is necessary. This article provides an introduction to the performance and interpretation of the single steps of a systematic review: defining a question, searching the literature, study selection, extracting information, assessing quality, and summarizing and presenting the results.

## Einleitung

Angesichts der Flut neuer Informationen ist es für Mediziner fast unmöglich, selbst im eigenen Spezialgebiet die Übersicht über den aktuellen Stand der Forschung zu erhalten. In dieser Situation erhalten systematische Übersichtsarbeiten (synonym: systematische Reviews) zunehmende Bedeutung. Übersichtsarbeiten werden dann als systematisch bezeichnet, wenn sie nach einem vorher festgelegten Studienprotokoll erstellt werden, in dem Ziele und methodischen Schritte soweit

möglich festgelegt sind. Schließen solche systematischen Übersichtsarbeiten eine zusammenfassende statistische Auswertung der behandelten Originalarbeiten ("primary studies") mit ein, werden sie als Meta-Analysen bezeichnet. Die Zahl von Meta-Analysen zu medizinischen Fragestellungen ist in den letzten Jahren sprunghaft angestiegen [1]. Die überwiegende Mehrzahl beschränkt sich auf die Erfassung und Bewertung randomisierter klinischer Studien, jedoch liegen auch Meta-Analysen anderer Studienformen in zunehmender Zahl vor.

<sup>1</sup> Projekt „Münchener Modell“, Ludwig-Maximilians-Universität

<sup>2</sup> Theoretische Chirurgie, Chirurgische Klinik und Poliklinik, Klinikum Innenstadt, Ludwig-Maximilians-Universität

Während im deutschen Sprachraum Meta-Analysen meist von Biometrikern durchgeführt werden, und die statistische Auswertung im Vordergrund steht, sind im englischen Sprachraum systematische Reviews bzw. Meta-Analysen häufig stärker im Sinne von tatsächlichen Übersichten aufgebaut; die klinische Relevanz und Interpretation stehen dort im Vordergrund.

Besondere Bedeutung dürften systematische Übersichtsarbeiten in den nächsten Jahren auch durch die Cochrane Collaboration erlangen [2,3]. Dieses weltweite Netzwerk hat zum Ziel, systematische Übersichtsarbeiten zu allen verfügbaren randomisierten klinischen Studien in allen Bereichen der Medizin zu erstellen, zu aktualisieren und zu verbreiten. Parallel dazu ist ausgehend von Kanada die "evidence-based medicine" Bewegung zu beobachten, deren Ziel es ist, klinische Entscheidungsprozesse primär auf der Basis der Ergebnisse randomisierter Studien aufzubauen [4].

Trotz ihrer wachsenden Bedeutung sind systematische Übersichtsarbeiten noch als junger Forschungsbereich mit erheblichem Entwicklungsbedarf zu beurteilen. Insbesondere Meta-Analysen stoßen keineswegs auf allgemeine Akzeptanz (siehe z.B. [5,6]). Hauptkritikpunkt ist, daß häufig methodisch und/oder klinisch nicht vergleichbare Studien zusammengeworfen werden. Dementsprechend ist zu fordern, daß den Ergebnissen von Meta-Analysen nicht blind vertraut wird, sondern daß sie sorgfältig hinterfragt und interpretiert werden.

Im folgenden soll eine Einführung in die methodischen Schritte von systematischen Übersichtsarbeiten und Meta-Analysen gegeben werden (weitergehende Literatur siehe [7-11])

## 1. Protokoll einer systematischen Übersichtsarbeit

Die Methodik der systematischer Übersichtsarbeit ähnelt der retrospektiven Auswertung von (prospektiv erhobenen) Krankendaten; dementsprechend sind systematische Übersichtsarbeiten aus methodischer Sicht gut geplanten prospektiven klinischen Studien unterlegen [5]. Um zumindest die Extraktion und Auswertung der vorhandenen Daten wie in einer guten retrospektiven Studie möglichst zuverlässig und systematisch durchzuführen, sind die einzelnen Arbeitsschritte in einem Protokoll vorab festzulegen. Wenn die Vorkenntnisse über das vorliegende Material zu gering sind, muß die existierende Literatur jedoch erst gesammelt und überblickt werden, um die Erstellung eines methodisch einwandfreien Protokolls überhaupt zu ermöglichen.

## 2. Fragestellung der Übersichtsarbeit

Wie jede wissenschaftliche Arbeit sollte auch eine systematische Übersichtsarbeit eine möglichst klare Fragestellung haben, z.B.: Ist die Mortalität nach perkutaner Koronarangioplastik (PTCA) bei Patienten mit

starker Angina pectoris in der Vorgeschichte ähnlich der nach Bypassoperation [12]? Die Fragestellung sollte bereits definieren, zu welchen Patientengruppen (z.B. Patienten mit Angina pectoris Anamnese), zu welchen Interventionen (PTCA, Bypassoperation) und zu welchem Zielkriterium (Mortalität) eine Aussage erarbeitet werden soll. Der Sinn solcher Übersichtsarbeiten ist im allgemeinen, die beobachteten Effekte zusammenzufassen, um die Aussagekraft und Entscheidungsrelevanz im Vergleich zu den einzelnen Studien zu erhöhen. Das Problem der oft ungenügenden Fallzahlen kann vermindert und unterschiedliche Standpunkte können einem Konsens zugeführt werden.

Systematische Übersichtsarbeiten können jedoch auch mit weniger eng definierten Fragestellungen durchgeführt werden, z.B.: Welche Evidenz liegt für die Wirksamkeit von Akupunktur bei chronischen Schmerzen vor? [13] Solch vage Fragestellungen können sinnvoll sein, wenn der allgemeine Wissensstand zum Thema noch gering bzw. unklar ist. Derartige systematische "state-of-the-art" Reviews haben aber mehr hypothesengenerierenden Charakter, und die Ergebnisse sind dementsprechend in Bezug auf die klinische Entscheidungsfindung mit größerer Zurückhaltung zu interpretieren.

## 3. Selektion relevanter Studien und Literatursuche

Ein guter systematischer Review sollte alle relevanten Arbeiten zu einem Thema einschließen, oder, falls dies nicht möglich sein sollte, eine möglichst repräsentative Stichprobe umfassen. Die Formulierung von Ein- und Ausschlusskriterien für die Auswahl der Originalarbeiten ist dabei ein zentraler Prozeß: Daher ist genau überprüfen, ob der angesetzte Filter tatsächlich die relevanten Studien selektiert. Bei Meta-Analysen werden entsprechend der Fragestellung Selektionskriterien auf folgenden Ebenen formuliert:

- \* Patienten (z.B. Patienten mit Major Depression entsprechend DSM-III-R Klassifikation),
- \* Interventionen (z.B. Maprotilin)
- \* Art der Kontrollgruppe (z.B. Placebo)
- \* Design/Maßnahmen zur Vermeidung von Verzerrungen (z.B. nur randomisierte, doppelblinde Studien)
- \* Zielkriterien (z.B. Hamilton-Depressionskala).

Häufig erfolgt auch eine Beschränkung auf englischsprachige Originalarbeiten, die in Medline geführt werden. Schließlich sind pragmatische Kriterien (z.B. ausreichende Präsentation der Studienergebnisse für eine biometrische Weiterauswertung) heranzuziehen, um zu entscheiden, ob die betreffende Originalarbeit tatsächlich in eine Auswertung miteinbezogen werden kann.

Da eindeutig belegt ist, daß durch Suchen in der größten und bekanntesten medizinischen Datenbank Medline zum Teil weniger als die Hälfte der tatsächlich existierenden relevanten Studien identifiziert werden [14,15], ist es unabdingbar, daß darüberhinausgehende Strategien zur Literatursuche Verwendung finden.

Durch den Einbezug der Datenbank Embase (die jedoch weniger häufig verfügbar und deutlich teurer ist) können beispielsweise viele nicht in Medline gelistete nichtenglischsprachige Publikationen erfaßt werden. Zusätzlich wünschenswert sind Kontakte mit Forschern im jeweiligen Bereich, Suchen in spezialisierten Datenbanken und das Prüfen der Bibliographien der bereits gefundenen Arbeiten mit dem Ziel, weitere Studien zur Fragestellung zu finden.

Gelegentlich wird argumentiert, daß eine Medlinesuche ausreiche, da damit in der Regel die qualitativ besten Arbeiten gefunden werden könnten; Artikel, die in nicht in Medline-gelisteten Zeitschriften erschienen sind, seien häufig nicht durch einen peer review gegangen und daher weniger zuverlässig. Dies trifft zwar zu, dennoch birgt ein solches Vorgehen ein erhebliches Risiko, relevante Studien zu übersehen. Einziges echtes Argument für eine Beschränkung auf Medline-Suchen sind beschränkte Ressourcen.

Effiziente Suchen in elektronischen Datenbanken wie Medline oder Embase erfordern einige Übung. Die Arbeiten zu einem bestimmten Thema sind oft nicht ohne weiteres durch eine einfache Suchstrategie auffindbar. Daher sollte die Suchstrategie in Zusammenarbeit mit sucherfahrenen Personen entwickelt und wiederholt überprüft werden.

Letztendlich ist es für den Leser einer systematischen Übersichtsarbeit nicht einfach zu überprüfen, ob die Suche tatsächlich adäquat war. Dies sollte der Autor einer systematischen Übersichtsarbeit berücksichtigen und dementsprechend die Nachvollziehbarkeit seines Vorgehens durch eine detaillierte Beschreibung gewährleisten.

Ein großes Problem von Übersichtsarbeiten ist die Möglichkeit eines "publication bias" [16,17]. Als "publication bias" bezeichnet man die selektive Publikation positiver Ergebnisse, während negative Ergebnisse unveröffentlicht bleiben. Liegt ein publication bias vor, so wird eine systematische Übersichtsarbeit publizierter Originalarbeiten ein zu positives Ergebnis erbringen. Daß "negative" Ergebnisse seltener als positive Ergebnisse publiziert werden, ist ein empirisch eindeutig belegtes Phänomen [16,18]. Das einzig sichere Gegenmittel wäre eine prospektive und transparente Registrierung aller klinischer Studien z.B. nach Erteilung eines positiven Ethikkommissionsvotums [19]. Derartige Register sind jedoch in den meisten Ländern nicht in Aussicht. Die einzige Möglichkeit für die Reviewer, mit diesem Problem umzugehen, ist, gezielt (z.B. durch persönliche Kontakte, Besuch von Tagungen etc.) nach unpublizierten Originalarbeiten bzw. Studienergebnissen zu suchen. Darüberhinaus gibt es statistische Techniken, die grobe Schätzungen bzw. Korrekturen erlauben (siehe Abschnitt Darstellung und Zusammenfassung der Ergebnisse). Alle diese Strategien können die Möglichkeit einer Verzerrung durch publication bias jedoch nicht vollkommen eliminieren. Die daraus resultierende Unsicherheit der Ergebnisse ist ein Nachteil; in der Publikation einer systematischen Übersichtsarbeit sollte das Problem auf die jeweilige Fragestellung bezogen immer diskutiert werden.

Durch eine vollständige und verzerrungsarme Literatursuche wird in aller Regel ein Pool von Arbeiten zusammengetragen, aus denen dann diejenigen selektiert werden müssen, die den Einschlusskriterien entsprechen. Nach Möglichkeit sollten mehrere, mindestens aber zwei unabhängige Beurteiler über den Einschluss der einzelnen Arbeiten entscheiden; die Häufigkeit von Nichtübereinstimmungen ist zu dokumentieren. Eine tabellarische Beschreibung der potentiell relevanten Studien, die ausgeschlossen wurden, sowie die Angabe der Ausschlußgründe sind wünschenswert. Die bibliographische Information aller gefundenen Originalarbeiten sollte jedoch in jedem Fall zugänglich sein (wenn dies zu umfangreich ist, kann ein Hinweis beigefügt werden, daß eine vollständige Literaturliste von den Autoren angefordert werden kann).

#### **4. Extraktion von Information aus den Originalarbeiten**

Sind die einzuschließenden Originalarbeiten identifiziert, müssen die Reviewer die relevanten Informationen aus diesen extrahieren. Nach Möglichkeit sollten hierbei standardisierte Methoden - in der Regel unter Verwendung von Formularen - angewendet werden. Die Formulare sollten an einer Anzahl ähnlicher Studien vorgetestet und gegebenenfalls optimiert werden. Vor allem bei der Extraktion der Ergebnisse sind oft erhebliche subjektive Entscheidungen durch den Reviewer zu treffen (z.B. Auswahl eines Zielkriterium, Dichotomisierung von Daten etc.), die in der Publikation der Übersichtsarbeit darzustellen und zu begründen sind. Nichtübereinstimmungen zwischen verschiedenen Reviewern sollten dokumentiert und diskutiert werden. Die Informationen werden nach Möglichkeit in einer Form extrahiert, die eine Weiterverarbeitung in einem statistischen Programm oder einer Datenbank erlaubt. Im Zweifelsfall sollten Informationen umfangreicher als unbedingt notwendig extrahiert werden, um zu vermeiden, daß bei späteren Zusatzfragen die Originalarbeiten nochmals bearbeitet werden müssen.

#### **5. Beurteilung der Qualität der Originalarbeiten**

Ein Haupteinwand gegen Meta-Analysen ist das "Zusammenwerfen" von qualitativ nicht gleichwertigen Studien [5,6]. Daher wird in vielen systematischen Übersichtsarbeiten zusätzlich (oder unter Verzicht auf eine Meta-Analyse) die "Qualität" der Originalarbeiten beurteilt. In aller Regel ist der Begriff Qualität dabei nicht näher definiert und wird von unterschiedlichen Autoren in völlig unterschiedlicher Weise ausgelegt [20].

Häufig wird Qualität synonym mit den Begriffen methodische Qualität und interne Validität verwendet. Eine kontrollierte Studie z.B. wird dann als "methodisch gut" oder "intern valide" bezeichnet, wenn die beobachtete Differenz zwischen den einzelnen

Gruppen mit hoher Wahrscheinlichkeit tatsächlich auf die unabhängige Variable zurückgeführt werden kann und Verfälschungen (Bias) weitgehend auszuschließen sind. Zentrale Kriterien hierfür sind eine adäquate Randomisation, Verblindung und eine möglichst alle aufgenommenen Patienten berücksichtigende Auswertung (intent to treat Analyse) [21].

Eine methodisch gute Studie kann klinisch jedoch völlig irrelevant oder gar irreführend und schlecht sein, z.B. wenn die Ergebnisse nur auf eine extrem selektierte Patientengruppe übertragbar sind, oder die Intervention in der untersuchten Form nicht den gängigen medizinischen Qualitätsstandards entspricht. In diesem Zusammenhang wird von der externen Validität einer Studie gesprochen [22].

Sinnvollerweise sollten in einer methodisch einwandfreien Qualitätsbeurteilung interne und externe Validität berücksichtigt werden. Besonders Aspekte der externen Validität sind aber schwer operationalisierbar und oft subjektiv. Zusätzlich kompliziert wird die Beurteilung der Qualität durch die Tatsache, daß dem Reviewer nur schriftliche Berichte, meist sogar nur komprimierte Publikationen vorliegen. Somit kann nur die Verfügbarkeit einer Information (z.B. wurde verblindet oder nicht) überprüft werden, nicht aber ob dieser Aspekt in der Studie tatsächlich adäquat durchgeführt und damit berücksichtigt wurde.

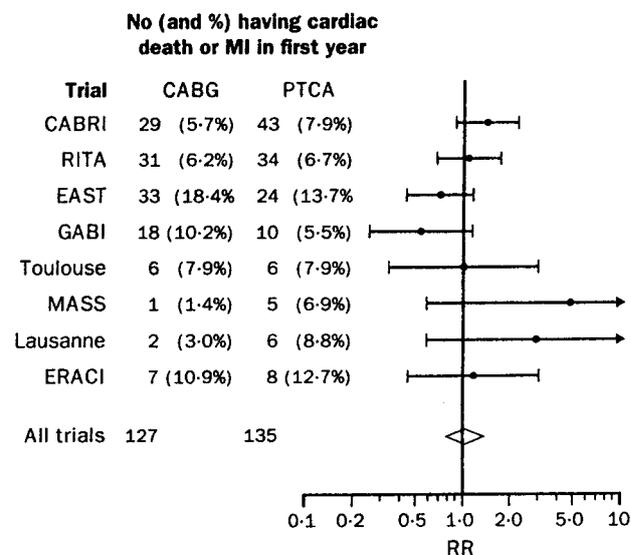
Angesichts dieser Schwierigkeiten sind Qualitätsbeurteilungen, die in vielen systematischen Übersichtsarbeiten in Form von Scores angewendet werden (Übersicht in [20]), bei denen für die Erfüllung einzelner "Qualitätskriterien" Punkte vergeben werden und ein Summenmaß gebildet wird, mit äußerster Vorsicht zu genießen [23]. Wendet man unterschiedliche Scoresysteme bei den gleichen Originalarbeiten an, ergeben sich zum Teil erheblich differierende Ergebnisse [24].

Aus unserer Sicht ist eine umfassende und klare Darstellung von Schlüsselmerkmalen zu Methodik und Beurteilung in Übersichtstabellen die beste Möglichkeit zur Handhabung dieses ungelösten Problems. Dabei sollte immer die Möglichkeit bestehen, auf ein bei der jeweiligen Originalarbeit besonders relevantes Problem in freiem Text hinzuweisen, da ein einziger schwerwiegender Fehler eine ganze klinische Studie in Frage stellen kann. Zusätzlich erscheint es durchaus sinnvoll, sowohl Scores als auch Einzelkriterien in einer Übersichtsarbeit zu benutzen, um überprüfen zu können, ob einzelne Qualitätsmerkmale einen Einfluß auf die Ergebnisse haben (Diese Methodik des multiplen methodischen Ansatzes ist in der Feldforschung als Triangulation bekannt [25]). So fanden z.B. Schulz et al., daß adäquat randomisierte und verblindete Studien weniger positive Ergebnisse erbrachten als die in diesen Punkten nicht adäquaten Studien [22].

In jedem Fall sind die in einer Übersichtsarbeit angewendeten Qualitätskriterien im Methodikteil klar und nachvollziehbar zu beschreiben und die Auswahl zu begründen.

## 6. Darstellung und Zusammenfassen der Ergebnisse der Originalarbeiten

Obwohl die methodischen Aspekte klinischer Studien äußerst wichtig sind, bleibt das Entscheidende das Ergebnis. Daher sollte in einem systematischen Review - soweit er nicht explizit auf methodische Fragestellungen begrenzt ist - die Darstellung und Zusammenfassung der Ergebnisse der Originalarbeiten im Vordergrund stehen. Klinisch interpretierbare quantitative Meta-Analysen (also integrierende statistische Auswertungen der Originalarbeiten) sind nur möglich, wenn eine ausreichende Zahl vergleichbarer Studien mit suffizienten Datenangaben (z.B. Mittelwert, Standardabweichungen, Gruppengröße) vorliegen. Eine typische Darstellung der Ergebnisse einer quantitativen Meta-Analyse zeigt die aus dem bereits oben erwähnten Beispiel zu PTCA und Bypassoperation entnommene Abbildung 1.



Legende zu Abbildung 1 (aus Pockock et al. [12])  
Beispiel zur Ergebnisdarstellung in Meta-Analysen.  
Todesfälle aufgrund kardialer Ursachen und Herzinfarkte bei Patienten nach perkutaner Angioplastik (PTCA) und Bypassoperation (CABG) im ersten Jahr nach Randomisation (Darstellung als relatives Risiko PTCA:CABG mit 95%-Konfidenzintervallen).

Die Ergebnisse der einzelnen Originalarbeiten (hier zur Mortalität innerhalb des ersten Jahres nach Intervention absolut und in Prozent) werden untereinander dargestellt. Für jede Studie wird ein Effektmaß (zentraler Punkt) mit dem zugehörigen 95%-Konfidenzintervall (Bereich um den zentralen Punkt) berechnet, im vorliegenden Beispiel das relative Risiko (Mortalitätsrate nach PTCA/Mortalitätsrate nach Bypassoperation). Ein relativ Risiko von 1 bedeutet dabei, daß kein Unterschied zwischen den beiden Behandlungsoptionen beobachtet wurde; ist das relative Risiko größer 1, ist die Mortalität nach PTCA höher als nach der Operation. Am unteren Ende der Abbildung findet sich schließlich das "gepoolte" Maß.

**Tabelle 1a:**

Effektmaße für dichotome (ja/nein) Daten

| Maß              | Erklärung                         | Vorteile/Nachteile   |
|------------------|-----------------------------------|--|
| Odds Ratio       | $\frac{a \times d}{b \times c}$   | statistisch robust/intuitiv schwer nachvollziehbar   |
| Relatives Risiko | $\frac{a / (a + b)}{c / (c + d)}$ | einfach nachvollziehbar/problematisch wenn extrem hohe oder niedrige Ereignisraten in der Kontrollgruppe |

a = Anzahl Patienten mit Ereignis in der Experimentalgruppe  
 b = Anzahl Patienten ohne Ereignis in der Experimentalgruppe  
 c = Anzahl Patienten mit Ereignis in der Kontrollgruppe  
 d = Anzahl Patienten ohne Ereignis in der Kontrollgruppe

**Tabelle 1b:**

Effektmaße für Meßdaten (z.B. Blutdruck in mmHg)

| Maß                                  | Erklärung                                      | Vorteile/Nachteile   |
|--------------------------------------|--|--|
| gewichtete Mittelwertsdifferenz      | $x_e - x_c$ gewichtet mit $1 / \text{Varianz}$ | einfach interpretierbar/nur verwendbar, wenn alle Daten auf der selben Skala gemessen wurden |
| standardisierte Mittelwertsdifferenz | $\frac{x_e - x_c}{sd_c}$                       | bei unterschiedlichsten Skalen anwendbar/intuitiv schwer nachvollziehbar                     |

$x_e$  = Mittelwert der Experimentalgruppe  
 $x_c$  = Mittelwert der Kontrollgruppe  
 $sd_c$  = Standardabweichung in der Kontrollgruppe

**Tabelle 2:**

Annahmen und Methoden zur Berechnung von gepoolten Effektmaßen (nach [8])

| Modellannahmen   | Methoden  | bei welchen Effektmaßen   |
|--|---|---|
| fixed effects (zugrundeliegende Originalarbeiten sind die einzigen von Interesse)                                    | Mantel-Haenszel<br>Peto<br>General variance based | Odds Ratio (auch relatives Risiko)<br>Odds Ratio<br>Odds Ratio, relatives Risiko, Differenzen |
| random effects (zugrundeliegende Originalarbeiten sind eine Zufallsstichprobe einer Population von Originalarbeiten) | Der Simonian-Laird                                | Odds Ratio, relatives Risiko, Differenzen   |

Diese Darstellung ist einfach und intuitiv nachvollziehbar, jedoch geht ihr eine ganze Reihe von Entscheidungen voraus. So ist wie in den Primärstudien selbst zu entscheiden, ob die Analyse "intent to treat" (also jeder

Patient, der ursprünglich zur PTCA-Gruppe zugeteilt wurde, wird analysiert, als habe er nur diese Therapie erhalten, auch wenn er später doch operiert wurde) oder "per protocol" durchgeführt wird. Ein Effektmaß ist zu

wählen (Beispiele siehe Tabelle 1) und zu entscheiden, auf der Basis welcher Annahmen und mit welcher Methode das gepoolte Effektmaß berechnet werden soll (siehe Tabelle 2).

Wichtig ist zu prüfen, ob die Ergebnisse der einzelnen Studien statistisch (häufig geprüft mit dem Chi-quadrat Test) und klinisch (nur subjektiv entscheidbar) homogen sind oder ob die Ergebnisse der einzelnen Studien sich überdeutlich unterscheiden.

Wenn aufgrund von Unterschieden bezüglich Design, Patienten, Interventionen, Zielkriterien oder Qualität eine quantitative Meta-Analyse nicht durchgeführt wird oder nicht durchführbar ist, müssen die einzelnen Studienergebnisse in Tabellen oder im Text beschrieben werden. Dies ist für Reviewer und Leser gleichermaßen mühsam. Keine gute Lösung ist die häufig in solchen Situationen angewendete Kategorisierung der Ergebnisse der Originalarbeiten (= vote counts), z.B. in "positive" und "negative" Ergebnisse. Derartige Kategorisierungen beruhen entweder darauf, ob die Ergebnisse statistisch signifikant waren bzw. nicht signifikant, oder auf den Schlussfolgerungen der Autoren der Originalarbeit oder der Reviewer. Im ersten Fall besteht, insbesondere wenn die Fallzahl in den Originalarbeiten zu gering war, ein erhebliches Risiko, klinisch relevante Unterschiede zu übersehen, im zweiten sind Nachvollziehbarkeit und Transparenz gering. Falls die in den Originalarbeiten präsentierten Daten ausreichend sind und zumindest die Zielkriterien vergleichbar sind, kann auf die graphisch-tabellarische Form von Abbildung 1 zurückgegriffen werden, ohne daß ein gepooltes Effektmaß gebildet wird. Diese prinzipiell äußerst sinnvoll Darstellung findet sich jedoch selten.

Grundsätzlich ist es wünschenswert, daß bei Meta-Analysen nicht nur die Effektmaße, die lediglich die Differenz zwischen den Vergleichsgruppen quantifizieren, angegeben werden, sondern auch die tatsächlich beobachteten Ergebnisse in den einzelnen Gruppen. Ein relatives Risiko von 0,5 ergibt sich z.B. sowohl in einer Studie mit einem Mortalitätsrisiko von 2/1000 in der Placebogruppe und 1/1000 in der Behandlungsgruppe wie auch in einer Studie mit Raten von 200/1000 bzw. 100/1000. Im ersten Fall müßten jedoch 1000 Patienten behandelt werden, um einen Patienten vor dem Tod zu bewahren (Maß bekannt als "number needed to treat"), im zweiten jedoch nur 10.

Ist die zentrale Analyse durchgeführt, werden in einer Meta-Analyse häufig Subgruppen- und Sensitivitätsanalysen angeschlossen. Subgruppenanalysen dienen wie in klinischen Studien dazu, zu überprüfen, ob bestimmte Patientenuntergruppen oder Interventionsbesonderheiten Auswirkungen auf das Effektmaß haben. Sensitivitätsanalysen werden durchgeführt, um zu überprüfen, wie "robust" das Resultat ist, d.h. ob zum Beispiel eine andere Art der Analyse (per protocol statt intention to treat, random effects statt fixed effects model) oder Design- und Qualitätsmerkmale der Originalarbeiten (z.B. Vergleich der Ergebnisse der doppelblinden Studien mit denen der nicht verblindeten) die Schlussfolgerungen verändern würden. Erstrebenswert ist unter

Umständen, Ursachen für eine Heterogenität der Ergebnisse der Originalarbeiten zu suchen (Methoden hierzu siehe [26]). Voraussetzungen für all diese Prüfungen ist jedoch eine ausreichende Anzahl verwertbarer Originalarbeiten.

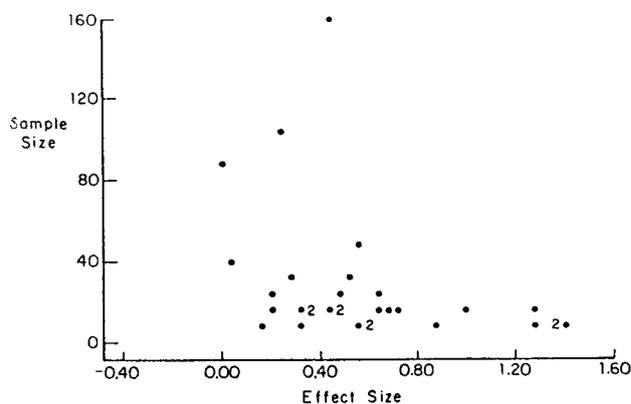


Abb. 2a

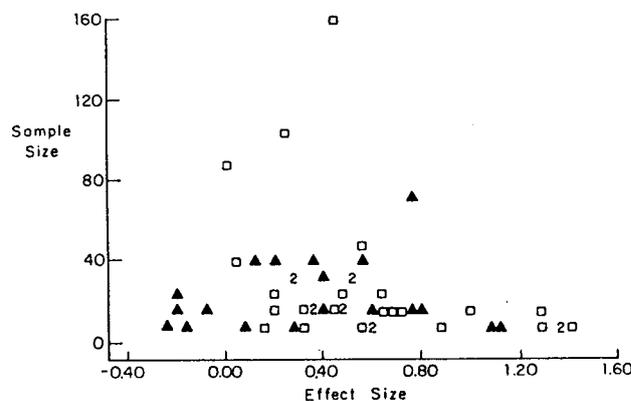


Abb. 2b

Legende zu den Abbildungen 2a und 2b (aus Light und Pillemer [9])

Beispiel für die Auswirkungen von publication bias. Funnel plots der publizierten (Abbildung 2a) sowie der publizierten  $\square$  und unpublizierten  $\blacktriangle$  Studien (Abbildung 2b) zu den Effekten eines psychologischen Trainingsprogrammes bei chirurgischen Patienten. Auf der y-Achse wird die Anzahl von Patienten in der jeweiligen Originalarbeit abgebildet, auf der x-Achse der beobachtete Effekt ([Mittelwert in der Experimentalgruppe - Mittelwert in der Kontrollgruppe] : Standardabweichung in der Kontrollgruppe). Vor allem Studien mit geringer Fallzahl und schwächeren bzw. negativen Effekten wurden nicht publiziert.

Dies gilt auch für die beste Strategie zur Abschätzung eines eventuellen publication bias, den "funnel plot" [9] (Trichter-Plot). Grundannahme für die Verwendung von funnel plots zu diesem Zweck ist, daß Studien mit großen Stichproben aufgrund des mit ihnen verbundenen Aufwandes, einem höheren Bekanntheitsgrad und meist auch ihrer höheren Qualität eine relativ große Wahrscheinlichkeit haben, unabhängig von ihrem Ergebnis publiziert zu werden. Dagegen ist bei kleineren Studien die Wahrscheinlichkeit publiziert zu werden geringer, wenn sie kein positives oder zumindest eindeutiges Ergebnis erbracht haben. In einem funnel plot werden die Effektmaße der einzelnen Studien gegen die Stichprobengröße aufgetragen. Wenn kein publication bias vorliegt (also die negativen und die positiven Studien unabhängig von ihrer Größe immer publiziert wurden), sollte die entstehende Abbildungen in etwa einem umgekehrten Trichter gleichen. Fehlt dagegen ein Teil des Trichters, muß mit dem Vorhandensein von publication bias gerechnet werden. Die Abbildungen 2a und b zeigen ein Beispiel aus dem Buch von Light und Pillemer [9], basierend auf Daten eines Reviews zur Wirksamkeit eines psychologischen Trainingsprogramms bei chirurgischen Patienten. Es wurden auch statistische Methoden vorgeschlagen, wie auf Basis des funnel plots das Ausmaß an publication bias und eine Näherung des wahren Effektmaßes geschätzt werden kann [27]. Wiederum muß aber gesagt werden, daß diese Methoden letztlich keine absolute Sicherheit bieten.

## 7. Präsentation

Der grundsätzliche Aufbau der Publikation einer Meta-Analyse ist der gleiche wie der jeder empirischen Studie (Einleitung, Methodik, Ergebnisse, Diskussion und Schlußfolgerungen). In Zeitschriftenpublikationen systematischer Übersichtsarbeiten können nur die wichtigsten Methoden und Ergebnisse beschrieben werden. Kliniker werden sich kaum für methodische Details interessieren, Biometriker und Epidemiologen dagegen eher. Aussagen zu Selektionskriterien, Suchmethoden, Bewertung und statistischen Methoden sind jedoch unabdingbar, um die Nachvollziehbarkeit bezüglich Zuverlässigkeit, Gültigkeit und Übertragbarkeit zu gewährleisten. Typischerweise enthalten systematische Übersichtsarbeiten eine Tabelle, in denen Patienten, Interventionen und Methodik der Originalarbeiten beschrieben werden und eine oder mehrere tabellarisch-graphische Darstellungen der Ergebnisse. Unter Umständen sind Tabellen zu ausgeschlossenen Studien und Qualitätsbewertungen sinnvoll.

## 8. Wann ist die Durchführung einer systematischen Übersichtsarbeit sinnvoll?

Die Erstellung systematischer Übersichtsarbeiten ist grundsätzlich dann sinnvoll, wenn die Fragestellung relevant ist, eine ausreichende Zahl von Originalarbei-

ten vorliegt und eine zuverlässige aktuelle Übersicht nicht verfügbar ist. Wenn eine Fragestellung bereits eindeutig geklärt und die zugrundeliegenden Informationen hinreichend bekannt sind, ist es jedoch fraglich, ob die Durchführung eines aufwendigen Reviews tatsächlich notwendig ist.

Meta-Analysen und systematische Übersichten für sich gesehen nicht aussagekräftiger Originalarbeiten sollten sicherlich nicht als Ersatz für die Durchführung adäquater prospektiver Studien dienen [4]. Ebenso wenig sinnvoll ist es, wenn Meta-Analysen dazu benutzt werden, irrelevante Effekte mit Hilfe großer Stichprobenumfänge statistisch signifikant zu machen [5].

Meta-Analysen wurden zu Beginn vor allem in der psychologischen, soziologischen und pädagogischen Forschung entwickelt und eingesetzt; in den letzten Jahren steigt ihre Zahl in der Medizin rasant. Systematische Übersichtsarbeiten eignen sich auch für die Ausbildungsforschung in der Medizin, beispielsweise zur Untersuchung der Effekte didaktischer Interventionen (siehe z.B. [28]).

## 9. Wer sollte eine systematische Übersichtsarbeit durchführen?

Der Aufwand der Durchführung sorgfältiger systematischer Übersichtsarbeiten variiert extrem nach Menge der zu referierenden Arbeiten, Vollständigkeit der publizierten Daten, Komplexität der Fragestellung usw. Er wird in aller Regel unterschätzt. Die Erstellung eines Protokolls, die Entwicklung und Erprobung der Extraktions- und Beurteilungsinstrumente, insbesondere die Literatursuche und die Diskussion von Nichtübereinstimmungen im Falle mehrerer Beurteiler, die Analyse und die Erarbeitung einer effektiven Präsentationsform sind zeitaufwendig und oft mühsam.

Die vielfältigen Arbeitsschritte lassen sich am besten in einem Team realisieren. Obwohl die optimale Zusammensetzung eines Review-Teams von der jeweiligen Fragestellung abhängt und im Alltag vor allem die verfügbaren Ressourcen entscheiden, scheinen folgende Bedingungen förderlich für eine Übersichtsarbeit: Alle relevanten Arbeitsschritte sollten von mindestens zwei, voneinander unabhängig extrahierenden und bewertenden Personen durchgeführt werden. Das Review-Team sollte über Fachkompetenz bezüglich des bearbeiteten Problems, der untersuchten Interventionen, der Methodik kontrollierter Studien und der Statistik verfügen. Manche Autoren sind der Meinung, daß eine fundierte Kenntnis von Indikation und Intervention nicht notwendig ist. Aus unserer Sicht sind zuverlässige fachbezogene Aussagen ohne derartige Kenntnisse nicht sinnvoll möglich; Reviews ohne entsprechend kompetente Beurteiler sollten sich daher auf formale Bewertungen beschränken und auf fachliche Interpretationen weitgehend verzichten.

Ein erhebliches Problem ist, daß Übersichtsarbeiten in Bereichen aktiver Forschung schnell veralten. Vom Abschluß der Analyse bis zum Erscheinen der Publikation in einer Zeitschrift vergeht meist mehr als ein Jahr. Reviewdatenbanken wie die Cochrane Database of

Systematic Reviews [3] werden regelmäßig aktualisiert und wenn nötig korrigiert. Elektronische Medien sind prinzipiell in Zukunft wahrscheinlich der bessere Weg zur Verbreitung von Reviews. Wer einen systematischen Review unternimmt, sollte dann aber auch bereit sein, diesen zu aktualisieren oder zumindest in einer Form durchzuführen, die eine Aktualisierung zuläßt, ohne daß alle Arbeitsschritte wiederholt werden müssen.

## Schlußbemerkung

Systematische Übersichten sind komplexe wissenschaftliche Arbeiten; der Forschungsbereich "Research Synthesis" ist noch jung und seine Methoden noch in Entwicklung oder noch nicht vollkommen ausgereift. Systematische Reviews können praktische Erfahrung und methodisch adäquate, prospektive Studien nicht ersetzen. Wenn sorgfältig durchgeführt und mit der angemessenen Zurückhaltung interpretiert, sind systematische Übersichtsarbeiten aber ein äußerst nützliches und letztlich unverzichtbares Werkzeug für Ärzte, andere Entscheidungsträger im Gesundheitsbereich und schließlich auch für Patienten sowie andere Wissenschaften, um mit der Flut neuer Forschungsergebnisse sinnvoll umgehen zu können.

## Literatur

1. Garfied E. Meta-analysis and the meta-morphosis of the scientific literature review. *Current Contents* 1991;October 28 (Number 43):5-9.
2. Godlee F. The Cochrane Collaboration. *Br Med J* 1994;309:969-970.
3. Cochrane Collaboration: The Cochrane Database of Systemic Reviews. London: BMJ Publishing Group, 1996.
4. Evidence-based medicine working group. Evidence-based medicine. *JAMA* 1992;268:2420-2425.
5. Feinstein A. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;48:71-79.
6. Eysenck HJ. Meta-analysis and its problems. *Br Med J* 1994;309:789-792.
7. Cooper H, Hedges LV (eds). *The handbook of research synthesis*. New York: Russel Sage Foundation, 1994.
8. Petitti DB. *Meta-analysis, decision analysis, and cost-effectiveness analysis*. New York: Oxford University Press, 1994.
9. Light RJ, Pillemer DB. *Summing up - the science of reviewing research*. Cambridge: Harvard University Press, 1984.
10. Chalmers I, Altman DG (eds). *Systematic reviews*. London: BMJ Publishing Group, 1995.
11. Spitzer WO et al. The Potsdam conference on meta-analysis. *J Clin Epidemiol* 1995;48 (no 1).
12. Pocock SJ, Henderson RA, Rickards AF et al. Meta-

analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet* 1995;346:1184-1189.

13. ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990;43:1191-1199.
14. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. *Contr Clin Trials* 1985;6:271-279.
15. Kleijnen J, Knipschild P. Searching the literature with on-line data-bases. *Pharmaceutisch Weekblad (Scientific Ed.)* 1992;14:316-320.
16. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Contr Clin Trials* 1987;8:343-353.
17. Kleijnen J, Knipschild P. Review articles and publication bias. *Arzneim-Forsch/Drug Research* 1992;42(I):587-591.
18. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;272:158-162.
19. Meinert CL. Towards prospective registration of clinical trials. *Contr Clin Trials* 1988;9:1-5.
20. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Contr Clin Trials* 1995;16:62-73.
21. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials *JAMA* 1995;273:408-412.
22. Hulley SB, Newman TB, Cummings SR. Getting started: the anatomy and physiology of research. In: Hulley SB, Cummings SR (eds). *Designing clinical research*. Baltimore: Williams & Wilkins, 1988.
23. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994;140:300-301.
24. Linde K, Worku F, Stör W, et al. Randomized clinical trials of acupuncture for asthma - a systematic review. *Forsch Komplementärmed* 1996;3 (im Druck).
25. Denzin NK. Triangulation. In: Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. Oxford, New York: Pergamon Press, 1990: 592-594.
26. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;309:1351-1355.
27. Hedges LV. Modeling publication selection effects in meta-analysis. *Stat Science* 1992;7:246-255.
28. Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. *JAMA* 1992;268:1111-1117.

## Anschrift des Verfassers

Klaus Linde  
Projekt "Münchener Modell",  
Ludwig-Maximilians-Universität, Kaiserstr. 9,  
D 80801 München