

# Systematische Übersichten und Meta-Analysen

## – Artikel Nr. 19 der Statistik-Serie in der DMW –

### Systematic reviews and meta-analyses

#### Autoren

A. Ziegler<sup>1</sup> S. Lange<sup>2</sup> R. Bender<sup>2</sup>

#### Institut

<sup>1</sup> Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Universität zu Lübeck

<sup>2</sup> Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

#### Schlüsselwörter

- ▶ Systematische Übersicht
- ▶ Meta-Analyse
- ▶ Heterogenität
- ▶ Zufällige Effekte

#### Key words

- ▶ Systematic reviews
- ▶ Meta-analysis
- ▶ Heterogeneity
- ▶ Random effects

#### Hintergrund



In verschiedenen medizinischen Gebieten gibt es eine große Zahl von Studien, die ein und derselben oder zumindest sehr ähnlichen Fragestellungen zu einer spezifischen Therapie nachgehen. Doch wie kann der klinisch tätige Arzt der immer schneller wachsenden Flut von Publikationen Herr werden? So sind allein im vergangenen Jahr mehr als 10 000 Einträge unter dem Schlagwort „randomized clinical trial“ in Medline [16] zu finden.

Darüber hinaus sind die Ergebnisse kontrollierter klinischer Studien häufig durch inkonsistente Ergebnisse geprägt. Das bedeutet, dass einige Studien den Nutzen einer Therapie zeigen können, während in anderen deren Unwirksamkeit oder gar eine schädliche Wirkung berichtet wird. Alternativ könnte es sein, dass die Fallzahlen der einzelnen Studien zu gering sind, um einen Effekt, der vielleicht gering ist, auch nachweisen zu können.

In den oben beschriebenen Fällen kann die Entscheidungsfindung über den Einsatz bzw. Nicht-Einsatz einer Therapie nur durch die gemeinsame Betrachtung und Analyse der gesamten Evidenz zu einer Fragestellung sowie deren qualifizierte Bewertung unterstützt werden. Diese Aufgabe übernehmen dabei systematische Übersichten (engl.: systematic reviews) und Meta-Analysen. Dieses Thema hat in den vergangenen Jahren erheblich an Bedeutung gewonnen, auch wenn die Idee schon 100 Jahre alt ist [21]. Die erste Meta-Analyse zur Untersuchung einer Intervention wurde bereits 1955 durchgeführt [2]; der Begriff selbst wurde allerdings erst im Jahr 1976 von Glass erstmalig verwendet [15]. Einen sehr schönen historischen Abriss zu diesem Themenkreis findet der interessierte Leser bei Egger und Smith [12]. Besondere Aspekte bei der Bewertung von Meta-Analysen aus Beobachtungsstudien beschreiben Blettner und Sauerbrei [4]. Richtlinien

für das Berichten von Meta-Analysen wurden von einer Reihe von Autoren und Zeitschriften gegeben [19; für eine Übersetzung des QUOROM-Statements: [www.cochrane.de/de/quorum.htm](http://www.cochrane.de/de/quorum.htm), 23, 25]. Empfehlenswert für einen tieferen Einstieg ist das Buch von Egger und Kollegen [13].

In systematischen Übersichten wird versucht, die Nachteile des klassischen Reviews zu vermeiden, indem methodische Standards sowohl auf die Identifikation und Selektion der in die Übersicht einzuschließenden Originalstudien als auch auf die Synthese deren Ergebnisse angewendet werden. Der Ablauf einer solchen Studie ähnelt dabei sehr dem üblichen Vorgehen bei einer klinischen Studie. So ist auch bei systematischen Übersichten das Erstellen eines Studienprotokolls erforderlich (s. z. B. Cochrane Reviewer Handbuch auf <http://www.cochrane.org/resources/handbook/>). Der gesamte Ablauf der Studie von der Formulierung der Fragestellung bis hin zur Interpretation bzw. Publikation des Endergebnisses wird als systematische Übersicht bezeichnet. Die statistische Zusammenfassung der Resultate der Einzelstudien wird Meta-Analyse genannt. Allerdings verwenden manche Autoren den Begriff Meta-Analyse und systematische Übersicht synonym (so z. B. in Ref. [9]).

#### Der Ablauf einer systematischen Übersicht



Im Folgenden werden die wesentlichen Komponenten einer systematischen Übersicht ausführlicher beschrieben. Zunächst sind, wie in kontrollierten klinischen Studien auch, die Ziele der Studie, die Hypothesen und eventuelle Subgruppenanalysen zu formulieren. Zur Sammlung und Auswertung von Daten im Rahmen einer systematischen Übersicht kann Software, wie z. B. der Review-Manager verwendet werden. Letzterer wird

#### Bibliografie

DOI 10.1055/s-2007-959042  
Dtsch Med Wochenschr 2007;  
132: e48–e52 · © Georg Thieme  
Verlag KG Stuttgart · New York ·  
ISSN 0012-0472

#### Korrespondenz

Professor Dr. rer. nat.

Andreas Ziegler

Institut für Medizinische  
Biometrie und Statistik,  
Universitätsklinikum Schleswig-  
Holstein, Campus Lübeck,  
Universität zu Lübeck  
Ratzeburger Allee 160, Haus 4  
23538 Lübeck  
eMail  
ziegler@imbs.uni-luebeck.de

von der Cochrane Collaboration kostenlos im Internet zur Verfügung gestellt (<http://www.cc-ims.net/RevMan>).

## Literaturrecherche

Ein wesentlicher Teil der Arbeit bei einem systematischen Review besteht darin, alle vorhandene Evidenz zusammenzutragen – unabhängig davon, ob die Ergebnisse der einzelnen Studien publiziert wurden oder nicht. Denn die Resultate nicht veröffentlichter Studien können systematisch von denen publizierter abweichen. Dieses wird sehr schön in der Arbeit von Egger und Smith illustriert [11]. Das dahinter stehende Problem besteht darin, dass Arbeiten mit einem Ergebnis, die die positive Wirkung einer Therapieform statistisch signifikant nachweisen, leichter und eher publiziert werden, als solche mit nicht signifikanten Resultaten [11]. Auf diese Weise kommt es sehr einfach zu einer systematischen Verzerrung in Richtung des positiven Effekts, der sog. „Publikationsverzerrung“ (engl.: publication bias) (für eine detaillierte Darstellung siehe z. B. Ref. [11]).

Eine wesentliche Quelle für die Literaturrecherche ist die Literaturdatenbank Medline [16]. Doch reicht diese Datenbank bei weitem nicht aus, da sie einen großen Anteil von Studien, primär aus dem europäischen Sprachraum, nicht enthält. Zur Identifikation von wichtigen Arbeiten, die nicht in Medline gelistet sind, kann die Suche mit der Excerpta Medica DataBase, kurz: EMBASE, erforderlich sein [26]. Für Therapiestudien empfiehlt sich die Datenbank der Cochrane Collaboration, das Cochrane Controlled Trial Register (CTR), die die vermutlich vollständigste Sammlung kontrollierter klinischer Studien enthält [24].

Neben der Suche nach Studien in Literaturdatenbanken ist es häufig erforderlich, die Referenzen bereits gefundener Publikationen oder von Übersichtsarbeiten nach weiteren Studien zu untersuchen, eine Handsuche in relevanten Zeitschriften und Kongressbänden vorzunehmen und auch auf dem jeweiligen Gebiet forschende Wissenschaftler und Institutionen nach nicht veröffentlichten Studien und deren Ergebnissen zu befragen. Wesentliches Qualitätskriterium einer systematischen Übersicht ist die transparente Darstellung sowohl der Literaturrecherche, z. B. anhand von klar vorgegebenen Schlagwörtern (engl. key words) in den Datenbanken, als auch der in die Meta-Analyse einfließenden Einzelstudien anhand von vorab definierten Ein- und Ausschlusskriterien.

## Datenextraktion und Beurteilung der Qualität der Studien

Bevor die Qualität der Studien beurteilt werden kann, müssen zunächst die relevanten Informationen aus den Artikeln extrahiert werden. Dieser Arbeit kommt später eine große Bedeutung zu, da die Ergebnisse durch die ausgewählten Informationen und den später zur Meta-Analyse verwendeten Studien beeinflusst werden können [12]. Auch aus diesem Grund sollte ein standardisierter Dokumentationsbogen zur Extrahierung der Daten verwendet werden. Empfehlenswert ist darüber hinaus zur Reduktion von Fehlern, dass zwei Personen unabhängig voneinander die Daten extrahieren [14]. Diese Arbeit wird durch das einheitliche Berichten über die Ergebnisse randomisierter (Therapie-) Studien bzw. diagnostischer Studien ganz erheblich vereinfacht (s. z. B. Empfehlungen in Ref. [1, 5]).

Trotz dieser Bemühungen zur Standardisierung lässt sich die Auswahl der Arbeiten vermutlich nicht vollständig objektivieren, da die in den Publikationen dargestellten Informationen häufig nicht alle Aspekte hinreichend abbilden. Daher empfehlen manche Autoren, die Einschlusskriterien für Studien so elementar wie möglich zu gestalten [14], um anschließend Sensitivitätsanalysen (siehe unten) durchzuführen. Dennoch erscheint es m. E. korrekter, dass bei systematischen Übersichten grundsätzlich nur solche Studien berücksichtigt werden sollten, die hohe Qualitätskriterien erfüllen, wie sie z. B. in den Anforderungen an eine patientenorientierte klinische Forschung beschrieben wurden [22].

## Zielkriterium und statistische Analyse

Wie in kontrollierten klinischen Studien müssen auch bei Meta-Analysen das Zielkriterium sowie das statistische Verfahren zur zusammenfassenden Analyse vorab ausgewählt und im Studienprotokoll niedergelegt werden. Üblicherweise wählt man für stetige Zielkriterien, wie z. B. Körpergewicht oder Blutdruck, den Mittelwertsunterschied zwischen Prüf- und Kontrollgruppe. Dabei wird der tatsächliche Unterschied zwischen den Gruppen durch den Grad der Erkrankung der eingeschlossenen Patienten beeinflusst. So könnte der Effekt einer Therapieform zur Gewichtsreduktion bei Personen mit extremer Adipositas zu einer größeren durchschnittlichen Gewichtsabnahme führen als bei Personen mit einer leichten Adipositas.

Bei binären Zielkriterien ohne Zensierung, wie z. B. Adipositas ja/nein oder Therapieerfolg ja/nein, können das Odds Ratio [3] aufgrund seiner guten statistischen Eigenschaften [7, 14] oder das relative Risiko gewählt werden: Mit beiden lassen sich die Ergebnisse der einzelnen Studien miteinander zu einem Gesamtergebnis kombinieren. Absolute Maße wie die absolute Risikoreduktion [3] oder das Effektmaß „number needed to treat“ (NNT) [3] haben den Vorteil der unmittelbaren klinischen Relevanz [14] und werden daher häufig bevorzugt. Allerdings sollte die Wahl des Effektmaßes von den grundsätzlichen Annahmen über den Therapieeffekt in Abhängigkeit von unterschiedlichen Baseline-Risiken geleitet werden [8].

Der nächste Schritt besteht in der statistischen Zusammenfassung der Ergebnisse der einzelnen Studien zu einem einzigen Resultat. Die einfache Mittelung der Effekte über die Studien wird dabei in der Regel nicht zu dem Ergebnis führen, das der Wahrheit am nächsten ist. Daher werden üblicherweise gewichtete Effektschätzer verwendet, um größeren Studien oder solchen mit einer geringeren Variabilität einen stärkeren Einfluss als kleineren Studien oder Studien mit größerer Variabilität zu geben. Die Analysetechniken selbst können im Wesentlichen in zwei Kategorien eingeteilt werden, den so genannten Modellen mit festen Effekten und den Modellen mit zufälligen Effekten [20]. Beim Modell mit festen Effekten geht man davon aus, dass sämtliche Studien ein und dieselben Ergebnisse lieferten, wenn sie denn nur unendlich groß wären. Diese Annahme ist sicherlich in der Regel unrealistisch, da schon kleine Unterschiede im Studiendesign, z. B. bei den Ein- und Ausschlusskriterien, zu verschiedenen Ergebnissen in dem Grad des Wirksamkeitsunterschieds zwischen zwei Therapieverfahren führen können. Bei Studien mit zufälligen Effekten wird hingegen diese zusätzliche Variabilität berücksichtigt, so dass hier die Konfidenzintervalle in der Regel etwas breiter sind als bei Modellen mit festen Effekten. Aufgrund der restriktiven

Annahmen der Modelle mit festen Effekten wird häufig die Verwendung von Modellen mit zufälligen Effekten empfohlen (siehe dazu auch [6, 20]). Innerhalb der Modellklasse mit zufälligen Effekten gibt es wiederum verschiedene Ansätze, die zu einer unterschiedlichen Gewichtung der Einzelstudien führen. Bekannte Methoden sind der Ansatz nach DerSimonian und Laird (DSL), die so genannte restricted maximum likelihood (REML) und Bayesianische Ansätze [6, 20]. Es ist hierbei zu berücksichtigen, dass sich für dieselben Daten in Abhängigkeit vom gewählten statistischen Modell unterschiedliche (zusammengefasste) Effektschätzer ergeben können.

### Sensitivitätsanalysen

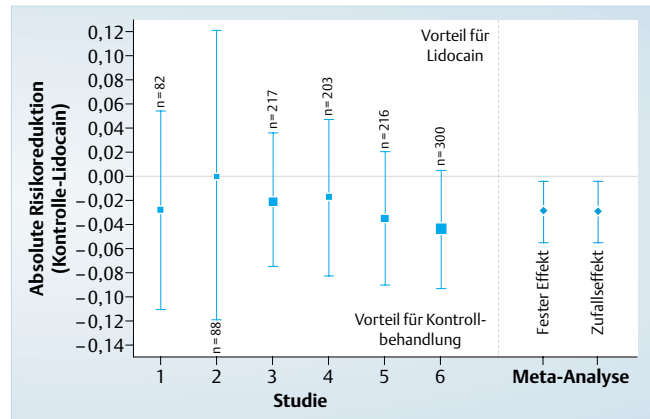
Welches statistische Verfahren zur Durchführung einer spezifischen Meta-Analyse am ehesten geeignet ist, wird kontrovers diskutiert. Daher ist die Durchführung von Sensitivitätsanalysen zur Überprüfung der Robustheit der Ergebnisse bei Verwendung verschiedener statistischer Verfahren dringend angeraten.

Die Ergebnisse der Meta-Analyse können sich auch bei Verwendung verschiedener sinnvoller Ein- bzw. Ausschlusskriterien der Studien (z. B. nur Einschluss von qualitativ hochwertigen Studien) voneinander unterscheiden, d. h., auch hierbei ist zu prüfen, ob die Ergebnisse robust gegenüber der Wahl der Studien sind. Dieses gilt insbesondere für Therapiestudien, die aufgrund signifikanter Ergebnisse vor Erreichen der ursprünglich geplanten Fallzahl abgebrochen werden, sowie für solche, bei denen keine Aussicht mehr existiert, noch einen Unterschied zu entdecken, und deshalb vorzeitig beendet werden.

Auch das Problem der Publikationsverzerrung lässt sich mit Sensitivitätsanalysen untersuchen. So können die Studien nach Studiengröße geschichtet werden. Wenn tatsächlich eine Publikationsverzerrung vorliegt, werden sich die Ergebnisse kleiner und großer Studien voneinander unterscheiden. Eine grafische Möglichkeit, eine eventuelle Publikationsverzerrung zu entdecken, bietet der so genannte „funnel plot“, bei dem die in den Einzelstudien beobachteten Effektschätzer (auf der x-Achse) gegen die jeweilige Fallzahl bzw. die in den Studien beobachtete Variabilität (auf der y-Achse) aufgetragen werden [10]. Liegt keine Publikationsverzerrung vor, werden sich die Einzelergebnisse aufgrund der zufälligen Variabilität zwischen den Studien symmetrisch und trichterförmig („funnel“) um den zusammengefassten Effektschätzer aus der Meta-Analyse verteilen.

### Heterogenität zwischen Studienergebnissen

Wenn die Resultate verschiedener Studien stark variieren, kann es sinnvoll sein, diese nicht zu einem einzelnen Ergebnis zusammenzufassen. Allerdings gibt es keine klaren Richtlinien, wann es noch sinnvoll ist bzw. wann nicht, die Studien zu kombinieren. Ein Ansatz ist, den Grad der Heterogenität der Studien zu untersuchen. So kann man z. B. testen, ob signifikante Unterschiede zwischen den Studien existieren. Der Nachteil dieser statistischen Tests ist, dass sie in der Regel nur eine geringe Macht besitzen, tatsächlich vorhandene Unterschiede zu entdecken. Daher ist eher die Anwendung von Heterogenitätsmaßen empfehlenswert, um den Grad der Heterogenität quantitativ zu beschreiben. Methoden hierzu wurden von Higgins und Thompson [17] entwickelt. Mög-



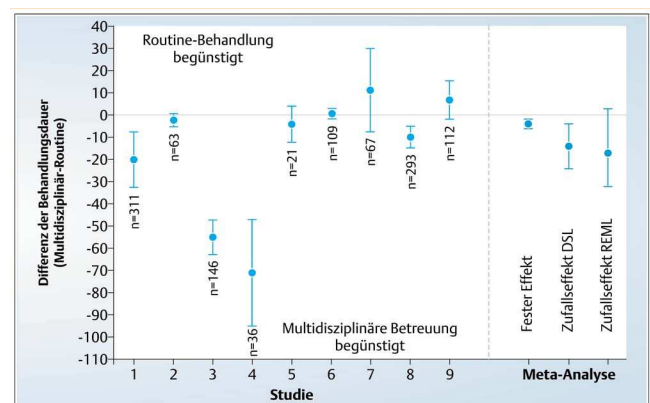
**Abb. 1** Wirksamkeit von Lidocain zur Reduktion von Mortalität im akuten Myokardinfarkt: x-Achse stellt die einzelnen Studien sowie die Meta-Analysen dar; y-Achse absolute Risikoreduktion (ARR). Für jede einzelne Studie und die Meta-Analysen sind das geschätzte ARR (Punkt) mit dem dazugehörigen 95% Konfidenzintervall (Schnurrbärte) dargestellt.

licherweise lassen sich Unterschiede auch schon mit einem forest plot (► **Abb. 1**) erkennen und inhaltlich erklären. In der Mitte sind die Punktschätzer der einzelnen Studien dargestellt, wobei die Fläche der Rechtecke das Gewicht der Einzelstudien symbolisiert. Für den zusammenfassenden Schätzer der Meta-Analyse wird ein Diamant benutzt. Die Schnurrbärte geben die Konfidenzintervalle an.

Bis vor wenigen Jahren wurden anstelle der forest plots Blobbogramme verwendet (s. z. B. [9]). Der Name Blobbogramm ist darauf zurückzuführen, dass dort ein Klümpchen in dieser Abbildung in der Mitte zu sehen ist (► **Abb. 2**), wobei diese für alle Stichproben gleich groß sind, also auf die Darstellung der Gewichtsinformation verzichtet wird.

### Beispiele

Zum Abschluss der Arbeit werden die unterschiedlichen Ergebnisse, die Meta-Analysen ergeben können, anhand zweier Beispiele aus der Literatur illustriert ([18], siehe auch [20, 24]).



**Abb. 2** Betreuung von Schlaganfallpatienten durch Team von Spezialisten mehrerer Disziplinen im Vergleich zum Routinemanagement: x-Achse stellt die einzelnen Studien sowie die Meta-Analysen dar; y-Achse Dauer des Krankenhausaufenthalts in Tagen (LOS). Für jede einzelne Studie und die Meta-Analysen sind die geschätzte LOS (Punkt) mit dem dazugehörigen 95% Konfidenzintervall (Schnurrbärte) dargestellt.

**Tab. 1** Untersuchung der Mortalität durch Prophylaxe mit Lidocain im akuten Myokardinfarkt (Quelle: Referenz [18], siehe auch [20]).

	Quelle	Anzahl randomisierter Patienten		Anzahl verstorbener Patienten	
		Lidocain	Kontrolle	Lidocain	Kontrolle
1	Chopra et al.	39	43	2	1
2	Mogensen	44	44	4	4
3	Pitt et al.	107	110	6	4
4	Darby et al.	103	100	7	5
5	Bennett et al.	110	106	7	3
6	O'Brian et al.	154	146	11	4
Gesamt		557	549	37	21

### Beispiel 1: Lidocain und Myokardinfarkt

Im ersten Beispiel wird untersucht, ob die Mortalität durch die Gabe von Lidocain beim akuten Myokardinfarkt gesenkt werden kann. Die Ergebnisse der Einzelstudien sind in **Tab. 1** dargestellt. Der forest plot (**Abb. 1**; mit Gewichten entsprechend dem Modell mit festen Effekten) zeigt die Risikounterschiede der sechs verschiedenen Studien. Das Rechteck in der Mitte gibt die beobachtete absolute Risikoreduktion (ARR) an. Die Schnurrbärte stehen für 95% Konfidenzintervalle.

Da alle Konfidenzintervalle die Nulllinie schneiden, ist keine einzelne der Studien in der Lage, einen positiven Behandlungseffekt von Lidocain statistisch signifikant nachzuweisen. Die Ergebnisse der sechs Studien sind insgesamt sehr homogen. Bis auf die zweite Studie, die keinen Unterschied zwischen Lidocain und der Kontrollbehandlung zeigt, deuten die fünf anderen Studien auf einen positiven Effekt von Lidocain. Mit der Meta-Analyse gelingt es jetzt, einen Vorteil von Lidocain gegenüber der Kontrollbehandlung nachzuweisen. Die absolute Reduktion der Mortalitätsrate liegt in der Größenordnung von 3% (95% Konfidenzintervall: 0,4% – 5,5%). **Abb. 1** zeigt, dass das verwendete statistische Modell mit festen Effekten und das Zufallseffektmodell zu identischen Ergebnissen und Schlussfolgerungen führen. In der Tat, die Ergebnisse und damit die Schlussfolgerungen sind auch für verschiedene spezifische Zufallseffektmodelle (DSL und REML) identisch.

### Beispiel 2: Betreuung von Patienten und Schlaganfall

In diesem Beispiel wird der Effekt einer Betreuung von Schlaganfallpatienten durch ein Behandlungsteam aus verschiedenen Dis-

ziplinen im Vergleich mit dem Routinemanagement von Schlaganfallpatienten untersucht [24]. Fallzahlen, Mittelwerte und Standardabweichungen der einzelnen Studiengruppen aus den neun betrachteten Studien sind in **Tab. 2** zusammengefasst. **Abb. 2** zeigt die Unterschiede der durchschnittlichen Liegedauer, d. h. Aufenthaltsdauer im Krankenhaus plus 95% Konfidenzintervall der verschiedenen Studien. Die zentrale Frage ist, ob durch den Einsatz eines multidisziplinären Behandlungsteams im Vergleich zum Routinemanagement die Aufenthaltsdauer reduziert werden kann. Vier Studien (1, 3, 4 und 8) zeigen einen Vorteil für die Spezialbetreuung. Anders als im ersten Beispiel sind hier die Ergebnisse der einzelnen Studien sehr heterogen (**Abb. 2**). Dieses Ergebnis wird auch durch einen statistischen Test auf Homogenität deutlich ( $p < 0,001$ ). Das Heterogenitätsmaß  $I^2$  [17] beträgt für diese Studien 96,65%. Das bedeutet, dass knapp 97% der Variabilität der Schätzer für den Behandlungseffekt auf die Heterogenität zwischen den Studien zurückzuführen ist. Daher haben wir bewusst in diesem Beispiel das Blobbogramm verwendet, um keine optische Verzerrung bei der Illustration der verschiedenen zusammenfassenden statistischen Verfahren zu erzeugen.

Der Leser möge beachten, dass durch die Verwendung einer stetigen Größe, d. h. der Aufenthaltsdauer in Tagen, die Studie mit der größten Fallzahl nicht automatisch die kleinsten Konfidenzintervalle aufweist. So liefern die Studien 2 und 6 die präzisesten Einzelergebnisse, obwohl die Studien 1 und 8 deutlich höhere Fallzahlen aufweisen. Diese unterschiedliche Variabilität der Ergebnisse zwischen den Studien ist in der Regel ein Zeichen für Heterogenität in der Patientenzusammensetzung der einzelnen Studien.

**Tab. 2** Vergleich der Betreuungsmodi von Schlaganfallpatienten in neun Studien: Team bestehend aus Spezialisten verschiedener Disziplinen im Vergleich mit Routinemanagement (Quelle: Referenz [18], siehe auch [20]).

	Quelle	Betreuung durch Spezialisten			Routinemanagement		
		N	Mittelwert LOS	StdAbw	N	Mittelwert LOS	StdAbw
1	Edinburgh	155	55,0	47,0	156	75,0	64,0
2	Orpington-Mild	31	27,0	7,0	32	29,0	4,0
3	Orpington-Moderate	75	64,0	17,0	71	119,0	29,0
4	Orpington-Severe	18	66,0	20,0	18	137,0	48,0
5	Montreal-Home	8	14,0	8,0	13	18,0	11,0
6	Montreal-Transfer	57	19,0	7,0	52	18,0	4,0
7	Newcastle 1993	34	52,0	45,0	33	41,0	34,0
8	Umea 1985	110	21,0	16,0	183	31,0	27,0
9	Uppsala 1982	60	30,0	27,0	52	23,0	20,0
Gesamt		548			610		

LOS: Dauer des Krankenhausaufenthalts (in Tagen)  
StdAbw: Standardabweichung

Sowohl das Blobbogramm (Abb. 2) als auch das Heterogenitätsmaß  $I^2$  und der statistische Test zeigen deutlich die Heterogenität der Studienergebnisse. Entsprechend ist die Anwendung eines Modells mit festen Effekten inadäquat. Diese Methode führt auch zu einem marginalen Effekt zugunsten der multidisziplinären Behandlung der Schlaganfallpatienten. Dabei ist zu beachten, dass dieses Ergebnis primär dadurch beeinflusst wird, dass die Studien 2 und 6, die beide keinen Effekt zugunsten der zu testenden Betreuung zeigen, einen großen Einfluss auf die Studienergebnisse erhalten: Ihnen wird in der Meta-Analyse mit festen Effekten das größte Gewicht beigemessen.

Anders fallen die Ergebnisse der Meta-Analysen mit zufälligen Effekten aus. Hier erhalten alle Studien tendenziell ähnliche Gewichte bei der Schätzung des gemeinsamen Behandlungseffekts. Entsprechend werden die Ergebnisse der Studien 1, 3 und 4 stärker berücksichtigt, was sich auch in dem Effekt über alle Studien hinweg zeigt. Die beiden dargestellten Meta-Analysen, die auf Modellen mit zufälligen Effekten basieren, gewichten die Studien selbst noch einmal unterschiedlich, so dass sich dadurch auch der Unterschied der Ergebnisse erklären lässt: Bei der sog. DSL-Methode werden hier die Studien weniger gleich gewichtet als bei dem REML-Verfahren. Das hat hier zur Folge, dass die DSL-Methode Signifikanz bei einer durchschnittlichen Verkürzung der Aufenthaltsdauer um 14 Tage (95% Konfidenzintervall -24 Tage bis -4 Tage) zeigt. Hingegen führt das REML-Verfahren zu einer durchschnittlichen Verkürzung der Aufenthaltsdauer um ca. 18 Tage (95% Konfidenzintervall -32 Tage bis +3 Tage), allerdings ohne statistische Signifikanz.

Insgesamt lässt sich für dieses Beispiel festhalten, dass die Betreuung von Schlaganfallpatienten im Durchschnitt zu einer verkürzten Aufenthaltsdauer im Krankenhaus führt. Jedoch sollte die Frage diskutiert werden, inwieweit eine Zusammenfassung der Studien bei den großen Unterschieden im Ergebnis in Abhängigkeit vom gewählten Modell sinnvoll ist. Weitere Untersuchungen wären notwendig, um herauszufinden, welche Aspekte der Spezialbetreuung zu der verkürzten Aufenthaltsdauer führen und sich auf andere Krankenhäuser übertragen lassen.

#### kurzgefasst

**Systematische Reviews vermeiden die Nachteile des klassischen Reviews, methodische Standards, wie sie für kontrollierte klinische Studien existieren, sowohl auf die Identifikation und Selektion der in die Übersicht einzuschließenden Originalstudien als auch auf die Synthese deren Ergebnisse angewendet werden.**

Dieser Beitrag ist eine überarbeitete Fassung aus dem Supplement Statistik aus dem Jahr 2004.

#### Literatur

- 1 Altman DG et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134: 663–694
- 2 Beecher HK. The powerful placebo. *JAMA* 1955; 159: 1602–1606
- 3 Bender R, Lange S. Die Vierfeldertafel. *Dtsch Med Wochenschr* 2007; 132: e12–e14
- 4 Blettner M, Sauerbrei W. Metaanalysen epidemiologischer Studien. *Med Klin* 1998; 93: 442–445
- 5 Bossuyt PM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Radiology* 2003; 226: 24–28
- 6 Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20: 825–840
- 7 Davies HTO, Crombie IK. What is meta-analysis? London: Hayward Medical Communications, 2001: 1–7
- 8 Deeks J. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; 21: 1575–1600
- 9 Döpfner S, Guggenmoos-Holzmann I. Meta-Analyse. *Dtsch Med Wochenschr* 1997; 122: 589–593
- 10 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997; 315: 629–634
- 11 Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998; 316: 61–66
- 12 Egger M, Smith GD. Meta-Analysis: Potentials and promise. *BMJ* 1997; 315: 1371–1374
- 13 Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care. Meta-Analysis in Context. London: BMJ Books, 2001
- 14 Egger M, Smith GD, Phillips AN. Meta-analysis: Principles and procedures. *BMJ* 1997; 315: 1533–1537
- 15 Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 5: 3–8
- 16 Greenhalgh T. How to read a paper: The Medline database. *BMJ* 1997; 315: 180–183
- 17 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539–1558
- 18 Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. *Arch Intern Med* 1989; 149: 2694–2698
- 19 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. *Lancet* 1999; 354: 1896–1900
- 20 Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; 18: 321–359
- 21 Pearson K. Report on certain enteric fever inoculation statistics. *BMJ* 1904; 3: 1243–1246
- 22 Schäfer H. Anforderungen an eine patientenorientierte klinisch-therapeutische Forschung. *Dtsch Med Wochenschr* 1997; 122: 1531–1536
- 23 Stroup DF et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA* 2000; 283: 2008–2012
- 24 The Cochrane Controlled Trial Register. Oxford: Update Software. Updated quarterly. The Cochrane Library, ed 2004
- 25 Weed DL. Methodologic guidelines for review papers. *J Nat Cancer Inst* 1997; 89: 6–7
- 26 Woods D, Trewheellar K. Medline and Embase complement each other in literature searches. *BMJ* 1998; 316: 1166