

Metaanalysen lesen und interpretieren: eine praktische Anleitung

A Practical Approach to Reading and Interpreting Meta-Analyses

Gesine Weckmann, Jean-François Chenot, Katrin C. Reber

Zusammenfassung: Ergebnisse klinischer Studien zu einer bestimmten Problemstellung werden in Metaanalysen zusammengefasst. Der vorliegende Artikel möchte eine praktische Anleitung bieten, Metaanalysen zu lesen und zu interpretieren. Die Ergebnisse von Metaanalysen werden oft in Grafiken und Diagrammen abgebildet, um einen schnellen Überblick der Ergebnisse zu ermöglichen. Häufig vorkommende grafische Darstellungen wie Forest-Plot und L'Abbé-Plot sowie das Heterogenitätsmaß I^2 werden in diesem Artikel erklärt.

Schlüsselwörter: Metaanalyse; Forest-Plot; L'Abbé-Plot; I^2

Summary: This paper aims to help family practitioners to quickly read and evaluate meta-analyses to keep up-to-date with current evidence-based treatment methods. Results of meta-analyses are often presented in plots and diagrams, to enable a quick review of the results. We explain how to interpret frequently used graphical displays like Forest plot and L'Abbé plot as well as heterogeneity measure I^2 .

Keywords: Meta-Analysis; Forest Plot; L'Abbé Plot; I^2

Einleitung

Eine wichtige Grundlage für Entscheidungen in der Patientenversorgung sind mit klinischen Studien evaluierte Diagnose- oder Behandlungsmethoden. Aufgrund der Fülle von Studien ist es dem einzelnen Arzt nicht möglich, alle für ein klinisches Problem relevanten und manchmal widersprüchlichen Studien zu identifizieren und zu bewerten. Daher werden Übersichtsartikel (Reviews) von Klinikern bevorzugt. Man unterscheidet **narrative** (unsystematische) **Reviews** von **systematischen Reviews**, die Ergebnisse aus Originalarbeiten zu einem Thema nach festgelegten Kriterien deskriptiv zusammenfassen. Eine Metaanalyse geht darüber hinaus und fasst die Ergebnisse mehrerer Einzelstudien mittels statistischer Methoden zusätzlich noch quantitativ zu einem globalen Ergebnis zusammen [1, 2]. Metaanalysen haben als Bestandteil systematischer Reviews stetig an Bedeutung gewonnen [3]. Ihnen wird die höchste Stufe an Evidenz zugeschrieben [4]. Ziel dieses Artikels ist ei-

ne praktische Anleitung zur Interpretation von Metaanalysen.

Systematische Reviews und Metaanalysen, wie sie z.B. von der Cochrane Collaboration [5] erstellt werden, erlauben einen raschen Überblick über die Studienlage und sind oft Grundlage für Leitlinienempfehlungen [6]. Metaanalysen ermöglichen es auch, Gründe für Inkonsistenzen und Widersprüche bei Einzelergebnissen zu explorieren. Die Ergebnisse systematischer Übersichtsarbeiten werden meist grafisch dargestellt, um sie schnell erfassen zu können. Bei der Interpretation ist zu beachten, dass die Ergebnisse der Metaanalyse nur so zuverlässig sind, wie die in sie eingeflossenen Einzelstudien [1]. Bei der Bewertung der Ergebnisse von systematischen Reviews und Metaanalysen sollte kritisch geprüft werden, ob die methodische Vorgehensweise bei der Auswahl und Zusammenfassung der Studien nachvollziehbar ist. Neben der methodischen Qualität muss auch immer die klinische Relevanz des gewählten Endpunkts bewertet werden. Ein Beispiel

für irrelevante Studienendpunkte sind Surrogatparameter wie HbA_{1c}-Senkung bei Diabetikern statt Endpunkte wie Senkung der kardiovaskulären Morbidität. Auch die ausgewählte Vergleichstherapie muss bei nicht-placebokontrollierten Studien kritisch geprüft werden.

Fragestellung und Studienauswahl

Eine präzise formulierte Fragestellung ist von zentraler Bedeutung. Sie bestimmt die Suchstrategie sowie Ein- und Ausschlusskriterien für die Auswahl der Einzelstudien und die zu extrahierenden Informationen für die Metaanalyse [7]. Sowohl für die klinische Frageformulierung als auch für die Forschungsfrage in der Metaanalyse hat sich das **PICO**-Schema bewährt (Tab. 1): **P**atient und/oder Problem (Patientenpopulation/Erkrankung); **I**ntervention; **C**omparison (Kontrollgruppe/Vergleichsintervention); **O**utcome (relevante Endpunkte/Zielgröße). Das

PICO-Schema beeinflusst die Literaturrecherche und die Auswahl der Einzelstudien. Allerdings beantworten die für eine Metaanalyse zur Verfügung stehenden Studien nicht immer exakt die vom Kliniker gestellte Frage.

Die genaue Definition der Patientenpopulation ist wichtig, weil z.B. Patienten aus Spezialambulanzen sich stark von denen in der Hausarztpraxis unterscheiden können. Diese haben möglicherweise schwerere Verläufe oder eine andere Ätiologie der Beschwerden und müssen deshalb ausgeschlossen oder separat analysiert werden [9]. Unterschiedliche Kriterien für den Studieneinschluss können zu abweichenden Ergebnissen führen. Die Vergleichsintervention muss nicht immer ein Placebo sein, sondern kann auch eine andere geeignete Alternative sein. Die Messung der Endpunkte muss am besten identisch sein, ansonsten muss sie konvertierbar sein. Für manche Fragestellungen ist auch der Zeitraum (**Timeframe**), in dem der Endpunkt beobachtet wird, entscheidend; dies wird dann als **PICOT** bezeichnet.

Wie präzise ist der errechnete Behandlungseffekt?

Der Unterschied im Studienendpunkt zwischen zwei Behandlungen wird als Effekt bezeichnet. Patienten und Ärzte möchten nicht nur wissen, ob eine Behandlung effektiv ist, sondern auch wie groß der Effekt ist. Ein häufig verwendetes Effektmaß für kontinuierliche Endpunkte ist die **gewichtete mittlere Differenz** (Weighted Mean Difference, WMD). Sie kann zur Beschreibung des Gesamteffekts in Metaanalysen gepoolt werden. Dabei erhalten die Einzelstudien ein unterschiedliches Gewicht, um die Präzision des Effektschätzers zu berücksichtigen. Üblicherweise fließt hier die Größe der Studie ein. Alternativ zur WMD können die Effekte auch als **relatives Risiko** (RR) oder kombinierte **Odds Ratio** (OR) dargestellt werden.

Der in kontrollierten Studien gefundene Effekt ist immer nur eine Schätzung mit einem **Konfidenzintervall**, in dem der wahre Effekt mit einer vorher bestimmten Wahrscheinlichkeit von meist 95 % liegt. Oft gibt es zu einem klinischen Prob-

Ein Arzt möchte sich zur Wirksamkeit einer lokalen Behandlung von akuten Halsschmerzen informieren. Es gibt kontrollierte Studien zu Einzelsubstanzen. Er findet auch eine Metaanalyse zur Schmerzinderung durch Lutschtabletten mit Ambroxol. Anhand dieser Metaanalyse wird das PICO-Schema beispielhaft erläutert [8].

PICO-Schema	Klinische Frage	Forschungsfrage in der Metaanalyse
Population/Problem	Patienten in der Hausarztpraxis mit akuten unkomplizierten Halsschmerzen	Kontrollierte Studien im ambulanten Setting wurden eingeschlossen. Studien zu Halsschmerzen nach Intubation und Studien zu Husteninderung wurden ausgeschlossen.
Intervention	Lokale Behandlung von Halsschmerzen	Lutschtabletten mit 20 und 30 mg Ambroxol (ohne Minzgeschmack).
Control	Placebo oder geeignete alternative Therapie	Vergleichstherapie waren Lutschtabletten ohne Wirkstoff mit Minzgeschmack. Das ist kein „reines“ Placebo.
Outcome	Schmerzinderung, Nebenwirkungen	Zusammenfassend wird berichtet: 1. die Schmerzreduktion 3 h nach Einnahme einer Lutschtablette als gewichtete mittlere Differenz zwischen Ambroxol und Placebo, 2. die Schmerzfreiheit in % nach 2 und 3 Tagen mit mehreren Lutschtabletten, 3. Nebenwirkungen.

Tabelle 1 Auswahl einer geeigneten Metaanalyse nach dem PICO-Schema

lem nur mehrere relativ kleine Studien mit großen Konfidenzintervallen und schwankenden Effekten. Es ist auch möglich, dass ein vorhandener Effekt im Zufallsrauschen nicht erkannt wird; dann spricht man davon, dass die **Power** der Studie zu klein ist. Bei größeren Studien ist die Wahrscheinlichkeit, dass Zufallsabweichungen das Ergebnis verfälschen, geringer. Studien zur gleichen Fragestellung und mit vergleichbaren Studienendpunkten können in einer Metaanalyse zusam-

mengefasst werden, um eine größere Power und eine statistisch präzisere Schätzung zu erreichen [1].

Wie konsistent sind die Einzelstudienresultate?

Bei der Interpretation der Metaanalyse sollte das Ausmaß, inwieweit sich die Ergebnisse der in die Metaanalyse eingeflossenen Studien unterscheiden (d.h. Heterogenität und Präzision), bewertet werden.

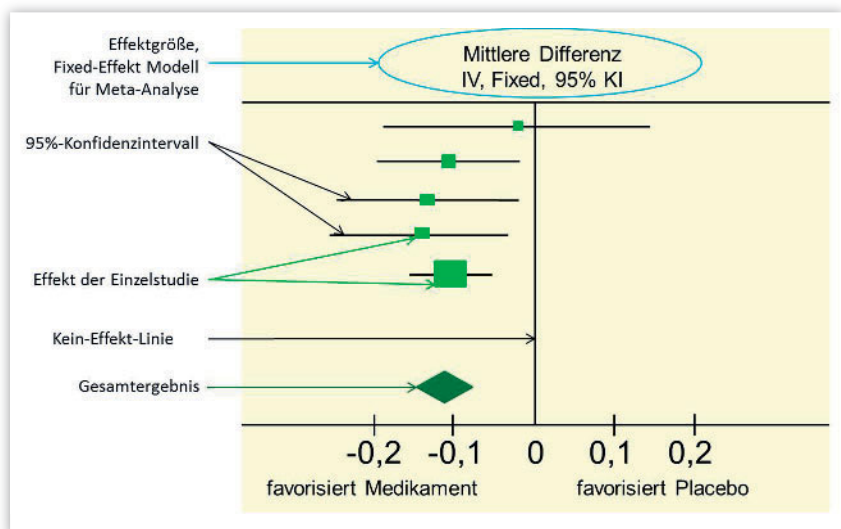


Abbildung 1 Beispiel Forest-Plot [8]

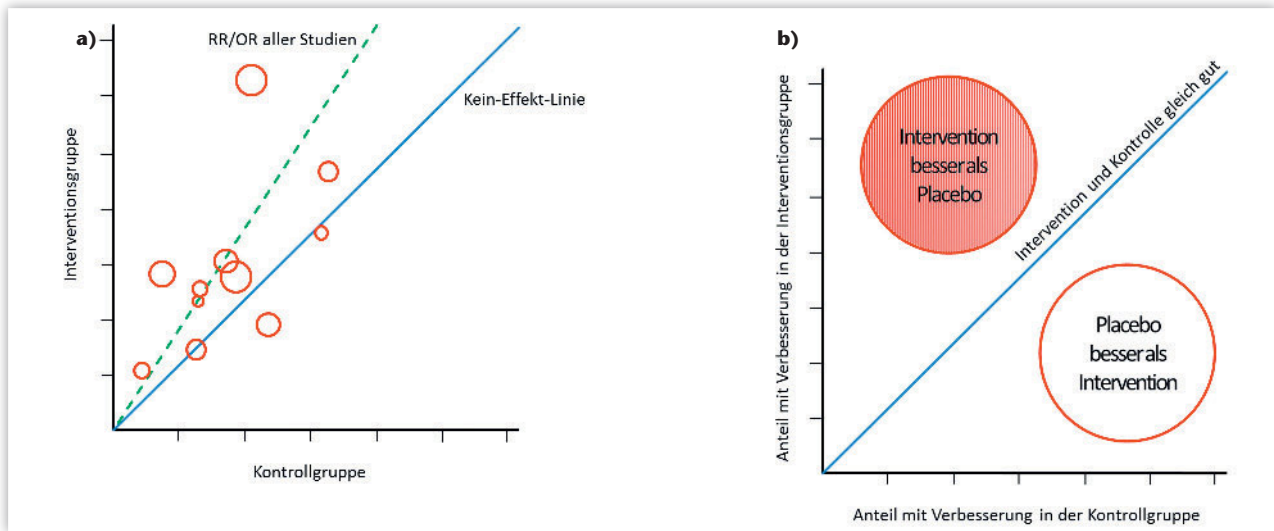


Abbildung 2 L'Abbé-Plot. a. Die blaue Linie stellt die Kein-Effekt-Linie (0-Linie) dar. Die gestrichelte grüne Linie stellt den Gesamteffekt aller Studien als RR oder OR dar. Die roten Kreise stellen die Ergebnisse der einzelnen Studien dar, wobei die Größe die Gewichtung der Studien repräsentiert. b. Schematische Darstellung eines L'Abbé-Plots [11, 12]

Grafische Darstellungen von Metaanalysen

Der **Forest-Plot** (Abb. 1) ist die am häufigsten verwendete grafische Methode, um die Effektgrößen aller in die Metaanalyse aufgenommenen Studien zusammen mit dem gepoolten Ergebnis (Gesamteffekt) darzustellen [10]. Der Therapie-Effekt wird als gewichtete mittlere Differenz im Vergleich zur Placebogruppe oder alternativen Therapie auf der horizontalen X-Achse abgetragen. Jede Studie wird durch ein Symbol dargestellt. Je größer die Studienpopulation, desto größer ist dieses Symbol. Die Symbole der Studienergebnisse sind um eine vertikale Linie dargestellt, die „keinen Effekt der Intervention“ repräsentiert: die Kein-Effekt-Linie (Abb. 1).

Durch die Symbole der einzelnen Studienergebnisse verläuft eine horizontale Linie, die das Konfidenzintervall darstellt. Überschneidet das Konfidenzintervall die Kein-Effekt-Linie, besteht kein (statistisch) signifikanter Unterschied zwischen Intervention und alternativer Therapie. Ein statistisch signifikanter Unterschied bedeutet aber nicht automatisch, dass dieser auch klinisch relevant ist. Die erste Studie (oben im Bild, Abb. 1) hat wenig Power und ein entsprechend großes Konfidenzintervall, das die Kein-Effekt-Linie überschneidet. Der beobachtete Effekt ist nicht signifikant. Das aus den Effekten der Einzelstudien ge-

poolte Gesamtergebnis der Metaanalyse wird als Raute dargestellt. Die Breite der Raute repräsentiert das Konfidenzintervall. Das gepoolte RR (oder die gepoolte OR) wird zusätzlich als gestrichelte Linie dargestellt.

Der L'Abbé-Plot (Abb. 2) ist eine andere Möglichkeit, Studienergebnisse zusammenfassend darzustellen [11–13]. Der L'Abbé-Plot wird verwandt bei Studien mit ja/nein-Ergebnissen, die man dichotome Endpunkte nennt. Ein Beispiel für einen dichotomen Endpunkt kann sein, ob der Patient mit dem Rauchen aufhört oder nicht. Im L'Abbé-Plot werden Studien wie folgt dargestellt: Für jede Studie werden die Ereignisrate (Risiko) in der Interventionsgruppe auf der Y-Achse und die Ereignisrate in der Kontrollgruppe auf der X-Achse dargestellt. Jeder Kreis repräsentiert eine individuelle Studie, wobei die Größe des Kreises proportional zur Studiengröße oder Gewichtung ist. Die 45°-Linie ist die Kein-Effekt-Linie. Studien mit identi-

scher Ereignisrate in der Interventions- und Kontrollgruppe liegen auf dieser Linie. Oberhalb der Kein-Effekt-Linie liegen Studien, die einen günstigen Effekt in der Interventionsgruppe aufweisen, unterhalb befinden sich Studien, die einen günstigen Effekt in der Kontrollgruppe zeigen. Der Unterschied der Ereignisraten zwischen interventions- und Kontrollarm in einer Studie ist umso größer, je weiter der Kreis von der Kein-Effekt-Linie entfernt liegt.

Heterogenität und Heterogenitätsmaß I²

Man kann die Effekte mehrerer Studien in einer Metaanalyse nur sinnvoll zusammenfassen, wenn die Einzelstudien ähnlich (homogen) sind. Es gibt aber immer Unterschiede (**Heterogenität**) zwischen Studien, bedingt durch unterschiedliche Patientenpopulationen, Interventionen oder Messmethoden („Äpfel und Birnen“-Problem). Große Unterschiede in den Effektgrößen oder Konfidenzintervallen deuten auf systematische Unterschiede zwischen den Studien hin. Das schränkt die Vertrauenswürdigkeit des gepoolten Gesamtergebnisses ein [14, 15].

Bei einer formalen Bewertung der Heterogenität soll geklärt werden, ob und in welchem Maße eine zufällige Variabilität für die Unterschiede verantwortlich ist [16]. Das am häufigsten ver-

I ² = 0 %	Keine Heterogenität
I ² = 25 %	Geringe Heterogenität
I ² = 50 %	Mittlere Heterogenität
I ² = 75 %	Hohe Heterogenität

Tabelle 2 Interpretation des I² = Heterogenitätsmaß [14]

Dr. med. Gesine Weckmann ...



... ist Ärztin in Weiterbildung.
Sie arbeitet an der Abteilung Allgemeinmedizin des Instituts für
Community Medicine der Universitätsmedizin Greifswald.

wendete Maß für Heterogenität ist I^2 [14]. Die Annahme hinter I^2 ist, dass die Variabilität der Studienergebnisse auf zwei Komponenten beruht: 1. auf systematischen Unterschieden zwischen den Studien und 2. auf zufälligen Abweichungen. I^2 misst den Anteil der Gesamtstreuung, der auf systematischen Unterschieden zwischen den Studien basiert und nicht durch Zufall entstanden ist. I^2 kann Werte zwischen 0 % und 100 % annehmen. Bei einem I^2 -Wert von 0 % wäre die Variabilität der Studienergebnisse vollständig auf Zufallschwankungen zurückzuführen. Bei einem Wert von 100 % wäre die Variabilität allein durch Unterschiede zwischen den Studien zu erklären. I^2 -Werte über 90 % sind selten [14, 17]. Wenn I^2 negativ ist, wird der Wert auf 0 % gesetzt. Eine Interpretationshilfe befindet sich in Tabelle 2. Als Vorteile von I^2 werden die einfache Interpretierbarkeit und die Unabhängigkeit von der Anzahl der in die Metaanalyse eingeschlossenen Studien genannt [18].

Unterschiedliche Qualität der Einzelstudien

Methodische Qualitätsunterschiede der Einzelstudien übertragen sich auf das metaanalytische Ergebnis (Garbage-in-Garbage-out-Problem) [1]. Metaanalysen enthalten eine Übersichtstabelle zur Bias der eingeschlossenen Studien. Das

am meisten genutzte Risk-of-Bias-Tool ist das der Cochrane Collaboration [19]. Das Ausschließen einzelner qualitativ schlechter Studien kann zu Selektionsbias führen. Eine mögliche Lösung ist eine Sensitivitätsanalyse, in der der gepoolte Effekt mit und ohne Einschluss der qualitativ schlechteren Studien analysiert wird.

Publikationsbias und fehlende Daten

Studien mit nicht-signifikanten oder negativen Effekten werden oftmals nicht publiziert und sind schwer zugänglich. Dadurch entsteht eine Verzerrung zugunsten positiver Ergebnisse. Bei nicht-signifikanten Ergebnissen werden relevante Kennzahlen, die für eine metaanalytische Zusammenfassung notwendig sind, häufig nicht mitgeteilt. Das Fehlen dieser kann zu einer Verzerrung des Gesamtergebnisses führen [18, 20].

Zur Abschätzung eines möglichen **Publikationsbias** eignet sich der Funnel-Plot [21]. Für jede Einzelstudie wird die Effektgröße (X-Achse) dem Standardfehler oder der Studiengröße (Y-Achse) gegenübergestellt. Eine sich hierbei ergebende asymmetrische Abbildung deutet auf einen Publikationsbias hin. Neben grafischen Verfahren werden auch statistische Testverfahren zur Abschätzung verwendet [22].

So wie es für klinische Studien „Leitlinien zur Berichterstattung“ gibt, gibt es für Metaanalysen das sogenannte **PRISMA-Statement** (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) [23]. Leser von Metaanalysen sollen darauf achten, dass die dort geforderten Kriterien eingehalten werden. Dazu wird oft eine Checkliste als Supplement veröffentlicht. Qualitativ gute Metaanalysen enthalten eine Übersichtstabelle zur Studienqualität der eingeschlossenen Studien.

Fazit

Metaanalysen sind eine gute Informationsquelle, um sich im Alltag mit überschaubarem Aufwand über die Effektivität einer Maßnahme zu informieren. Dennoch sollten die Ergebnisse einer Metaanalyse nicht unkritisch akzeptiert werden, sondern im Rahmen ihrer Übertragbarkeit in die Praxis und ihrer Limitationen bewertet werden.

Interessenkonflikte: Jean-François Chenot hat Honorare aufgrund einer Beratertätigkeit von WidO, Elsevier Health Risk, AOK erhalten; für Aussagen als Experte hat er Honorare des MDK bekommen. Die weiteren Autoren haben keine Interessenkonflikte angegeben.

Korrespondenzadresse

Dr. med. Gesine Weckmann
Abteilung Allgemeinmedizin
Institut für Community Medicine
Universitätsmedizin Greifswald
Fleischmannstraße 42-44
17475 Greifswald
Tel.: 03834 8622282
allgemeinmedizin@uni-greifswald.de

Literatur

1. Bown MJ, Sutton AJ. Quality control in systematic reviews and meta-analyses. *Eur J Vasc Endovasc Surg* 2010; 40: 669-677
2. Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 3. Ed. Edinburgh: Elsevier Churchill Livingstone, 2005: 281
3. Sutton AJ, Higgins JP. Recent developments in meta-analysis. *Stat Med* 2008; 27: 625-650
4. Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med* 2003; 96: 17-22
5. <http://www.cochranelibrary.com/> (letzter Zugriff am 19.05.2015)
6. Egger M, Smith GD, Altman DG. Systematic reviews in health care. Meta-analysis in context. London: BMJ Publications, 2001
7. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127: 380-387

8. Chenot JF, Weber P, Friede T. Efficacy of ambroxol lozenges for pharyngitis: a meta-analysis. *BMC Fam Pract* 2014; 15: 45
9. Green LA, Fryer Jr GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. *N Engl J Med* 2001; 344: 2021–2025
10. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322: 1479–1480
11. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; 107: 224–233
12. <http://www.medicine.ox.ac.uk/bandler/booth/glossary/labbe.html> (letzter Zugriff am 19.05.2015)
13. Higgins J. Considerations and recommendations for figures in Cochrane reviews: graphs of statistical data. *Cochrane Statistical Methods Group*, 2003 (update 2008)
14. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560
15. Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950; 37: 256–66
16. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract* 2008; 14: 951–957
17. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in Stata. In: Egger M, Davey Smith G, Altman DG (eds.). *Systematic reviews in health care: meta-analysis in context*, 4. Aufl. London: BMJ Publications, 2001: 347–369
18. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1538–1558
19. <http://ohg.cochrane.org/sites/ohg.cochrane.org/files/uploads/Risk%20of%20bias%20assessment%20tool.pdf>
20. Pigott TD. Methods for missing data in research synthesis. In Cooper H, Hedges LV (Eds.). *Handbook for research synthesis*. New York: Russell Sage, 1994: 163–175
21. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; 343: d4002
22. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629–634
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Med* 2009; 6: e1000097



DEGAM-Leitlinien frei im Netz

Die Leitlinien der Deutschen Gesellschaft für Allgemeinmedizin und Familienmedizin (DEGAM) stehen frei im Internet zur Verfügung. Die wissenschaftlich fundierten und vor der Veröffentlichung in Praxen erprobten DEGAM-Leitlinien richten sich nicht nur an Hausärzte, sondern auch an Patienten und Praxismitarbeiter. Neben der Langversion gibt es eine Kurzfassung als laminierte, zweiseitige Tischkarte im A5-Format. Mehrere tausend Leitlinien-Sets werden in Praxen und Universitäten in der täglichen Arbeit mit Patienten eingesetzt. Alle Module können auf der DEGAM-Leitlinien-Homepage (www.degam-leitlinien.de) oder auf der Homepage der AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, <http://leitlinien.net/>) bei Bedarf heruntergeladen und ausgedruckt werden.

Kontakt:

Philipp Leson
DEGAM-Bundesgeschäftsstelle
Friedrichstraße 133
10117 Berlin
Tel.: 030 209669800
Fax: 030 209669899
E-Mail: presse@degam.de
Homepage: www.degam.de

Dr. med. Anne Barzel
DEGAM-Geschäftsstelle Leitlinien
c/o Institut für Allgemeinmedizin
Universitätsklinikum Hamburg-Eppendorf
Martinistraße 52
20246 Hamburg
Tel.: 040 741059769
Fax: 040 741053681
E-Mail: leitlinien@degam.de